Dissertation

VTT PUBLICATIONS 754

Sanni Voutilainen

Fungal thermostable cellobiohydrolases

Characterization and protein engineering studies

VTT PUBLICATIONS 754

Fungal thermostable cellobiohydrolases

Characterization and protein engineering studies

Sanni Voutilainen

The dissertation will take place on 4th February 2011 at 12 (University of Jyväskylä, Ylistö Campus, lecture hall YAA303).

ISBN 978-951-38-7425-4 (soft back ed.)
ISSN 1235-0621 (soft back ed.)

ISBN 978-951-38-7426-1 (URL: http://www.vtt.fi/publications/index.jsp)
ISSN 1455-0849 (URL: http://www.vtt.fi/publications/index.jsp)

JULKAISIJA – UTGIVARE – PUBLISHER

VTT, Vuorimiehentie 5, PL 1000, 02044 VTT
puh. vaihde 020 722 111, faksi 020 722 4374

VTT, Bergsmansvägen 5, PB 1000, 02044 VTT
tel. växel 020 722 111, fax 020 722 4374

VTT Technical Research Centre of Finland, Vuorimiehentie 5, P.O. Box 1000, FI-02044 VTT, Finland
phone internat. +358 20 722 111, fax + 358 20 722 4374

Technical editing Leena Ukskoski

Edita Prima Oy, Helsinki 2010

Sanni Voutilainen. Fungal thermostable cellobiohydrolases. Characterization and protein engineering studies. Espoo 2010. VTT Publications 754. 98 p. + app. 55 p.

Keywords cellulase, protein engineering, thermostability, heterologous expression, *Saccharomyces cerevisiae*, high-throughput screening, random mutagenesis, site-directed mutagenesis, disulphide bridge

Abstract

Cellulases are important industrial enzymes, and they are already today utilized for example in the pulp and paper and textile industries. Due to their application potential in so-called second generation bioethanol production, starting from lignocellulosic raw materials, cellulases are currently in the centre of attention. Especially cellobiohydrolases are the key enzymes in total hydrolysis of biomass based processes. The currently known cellobiohydrolases from GH family 7 are, however, not highly active under industrial conditions.

In this work two different approaches were taken to improve the hydrolysis of crystalline cellulose. Firstly, thermotolerance and activity of two known fungal cellobiohydrolases from GH7 family was improved by protein engineering and secondly novel cellobiohydrolases were studied.

Engineering of GH7 cellobiohydrolases has been hindered because of difficulties in heterologous expression in a bacterial or yeast host. In this study a functional yeast expression system was developed for two single-module cellobiohydrolases. This heterologous expression system enabled protein engineering by random mutagenesis and site-directed mutagenesis. Random mutagenesis, carried out with error-prone PCR and followed by functional screening with an automated, thermostability screening method, was used for improving the thermostability of *Melanocarpus albomyces* Cel7B (*Ma* Cel7B). The stability and activity of *Ma* Cel7B was further improved through structure-guided protein engineering by introducing an additional disulphide bridge and a carbohydrate binding module. Rational mutagenesis was also used for engineering *Talaromyces emersonii* Cel7A (*Te* Cel7A). Altogether five individual S-S bridges were introduced to improve the stability of *Te* Cel7A. Three out of these five single S-S mutants had a clearly improved thermostability. These mutations were combined in a triple mutant, which had significantly improved

unfolding temperature (by 9°C) and ability to hydrolyse microcrystalline cellulose at 80°C.

In addition to the mutagenesis studies of known cellobiohydrolases, we found novel GH-7 family cellobiohydrolases, which have high activity on insoluble polymeric substrates. Three family 7 cellobiohydrolases originating from thermophilic fungi *Acremonium thermophilum*, *Thermoascus aurantiacus* and *Chaetomium thermophilum* were characterized and compared to the widely studied *Trichoderma reesei* Cel7A enzyme. The comparison revealed that all these novel cellobiohydrolases are promising for application purposes, because they were more thermostable than *T. reesei* Cel7A and more active in the hydrolysis of microcrystalline cellulose (Avicel) at 45°C.

Preface

This work was carried out at VTT Technical Research Centre of Finland in the Protein Engineering team and in the Development of Industrial Enzymes team. Former Vice President R&D Biotechnology, Prof. Juha Ahvenainen, Vice President R&D Bio- and process technology, Prof. Anu Kaukovirta-Norja and technology managers Niklas von Weymarn and Richard Fagerström are thanked for providing excellent working facilities and supporting the finalization of this thesis. I am grateful to Dr. Anu Koivula and Prof. Kristiina Kruus for leading the research teams. Financial support from the Academy of Finland, EU (project number NNE5-2001-00447, contract: ENK6-CT-2002-00604), Finnish Glycoscience Graduate School and VTT is gratefully acknowledged. Dr. Tapani Reinikainen is acknowledged for his valuable contribution to the work through the academy funding.

I would like to express my deepest gratitude to my supervisor Dr. Anu Koivula for her advice, encouragement and practical help during this work. I also wish to thank my co-authors Harry Boer, Markus Linder, Terhi Puranen, Juha Rouvinen, Jari Vehmaanperä, Janne Jänis, Patrick Murray, Maria Tuohy, Matti Siika-aho, Arja Lappalainen, Marika Alapuranen, Jarno Kallio, Satu Hooman and Liisa Viikari for their contribution to this work. I am especially grateful to Jari Vehmaanperä for his involvement and interest towards my work, and his patience when familiarizing me with the field of molecular biology at the very beginning of this route.

Professors Christian Oker-Blom and Olli Pentikäinen from University of Jyväskylä are acknowledged for supporting this thesis. Docent Tuomo Glumoff and Prof. Jerry Ståhlberg are thanked for the careful pre-examination of the thesis and their valuable comments to improve it. Michael Bailey is thanked for revising the language of this thesis.

I am indebted to all the people from the different teams in the lab; I would not have been able to do this without their support. I am especially grateful to Arja Kiema and Riitta Suihkonen for their assistance.

I thank all the team members of Development of Industrial Enzymes team and the former Protein Engineering team for help and support over the years. I am especially grateful to my former colleagues Sanna and Mamme for their friendship.

My warmest thanks go to my parents and friends for their support. Most of all I thank Harry for his encouragement and sharing his life with me.

Contents

Publications I–IV are not included in the PDF version.
Please order the printed version to get the complete publication
(http://www.vtt.fi/publications/index.jsp).

List of publications

This thesis is a summary of the following original articles, which are referred to as Publications I–IV. The articles are included as appendices in the printed version of the thesis.

I Voutilainen, S.P., Boer, H., Linder, M.B., Puranen, T., Rouvinen, J., Vehmaanperä, J. and Koivula, A. 2007. Heterologous expression of *Melanocarpus albomyces* cellobiohydrolase Cel7B, and random mutagenesis to improve its thermostability. Enzyme Microb Technol, Vol. 41, pp. 234–243.

II Voutilainen, S.P., Boer, H., Alapuranen, M., Jänis, J., Vehmaanperä, J. and Koivula, A. 2009. Improving the thermostability and activity of *Melanocarpus albomyces* cellobiohydrolase Cel7B. Appl Microbiol Biotechnol, Vol. 83, No. 2, pp. 261–272.

III Voutilainen, S.P., Murray, P.G., Tuohy, M.G. and Koivula, A. 2009. Expression of *Talaromyces emersonii* cellobiohydrolase Cel7A in *Saccharomyces cerevisiae* and rational mutagenesis to improve its thermostability Protein Eng Des Sel, Vol. 23, pp. 69–79.

IV Voutilainen, S.P., Puranen, T., Siika-Aho, M., Lappalainen, A., Alapuranen, M., Kallio, J., Hooman, S., Viikari, L., Vehmaanperä, J. and Koivula, A. 2008. Cloning, expression, and characterization of novel thermostable family 7 cellobiohydrolases. Biotechnol Bioeng, Vol. 101, No. 3, pp. 515–28.

The author's contribution to the appended publications

Publication I

The author planned the work together with the co-authors. The author conducted the laboratory work, except the synthesis of the cDNA and the bioreactor cultivations. The author interpreted the data, except the structural consequences of the mutations. The author had the main responsibility for writing the publication.

Publication II

The author planned the work together with Dr. Anu Koivula. The author conducted the laboratory work, except the cloning and *T. reesei* production of the CBM fusion protein. The author had the main responsibility of interpreting the data and writing the publication.

Publication III

The author planned the work together with Dr. Patrick Murray and Dr. Anu Koivula. The author planned and conducted the laboratory work and interpreted the data. The author had the main responsibility of writing the article.

Publication IV

The author had the responsibility of planning and conducting the laboratory work, except the cloning, production and purification of the enzymes. The author interpreted the data and had the main responsibility for writing the article.

List of abbreviations

At Cel7A	*Acremonium thermophilum* cellobiohydrolase Cel7A
BCA	2,2´-bicinchroninate
BMCC	bacterial microcrystalline cellulose
CBH	cellobiohydrolase
CBM	carbohydrate-binding module
CD	circular dichroism
CMC	carboxymethyl cellulose
CNPLac	2-Chloro-4-nitrophenyl-β-D-lactoside
Ct Cel7A	*Chaetomium thermophilum* cellobiohydrolase Cel7A
DP	degree of polymerization
DS	degree of substitution
DTT	dithiothreitol
EG	endoglucanase
FnIII	fibronectin typeIII
GH	glycoside hydrolase
Glc_2	cellobiose

Ma Cel7B	*Melanocarpus albomyces* cellobiohydrolase Cel7B
MALDI-TOF	matrix-assisted laser desorption/ionization/time of flight
MU	4-Methylumbelliferone
MULac	4-Methylumbelliferyl-β-D-lactoside
PAHBAH	p-hydroxylbenzoic acid hydrazine
PASC	phosphoric acid-swollen cellulose
SC	synthetic complete
Ta Cel7A	*Thermoascus aurantiacus* cellobiohydrolase Cel7A
T_m	melting temperature
TPI	triose phosphate isomerise
Tr Cel7A	*Trichoderma reesei* cellobiohydrolase Cel7A
Ura	uracil
wt	wild-type

1. Introduction

The action of microbial cellulolytic enzymes plays a crucial role in the carbon cycle on earth as they reduce the most common organic polymer, cellulose, to an exploitable form, glucose. Cellulases also have industrial importance due to their many applications for example in the pulp and paper and textile industries. Although these applications are economically important, the industrial importance of cellulases could rise to a new level if the utilisation of cellulose as a raw material for biofuels were to emerge and become economically feasible. The enzymatic production of fermentable sugars from cellulosic raw materials has only recently been initiated in full industrial scale. The use of enzymes instead of chemicals in cellulose hydrolysis supports sustainable development, but there is need for improvement of the efficiency of commercially available enzymes.

This work describes protein engineering studies to improve the thermostability and activity of fungal Cel7 cellobiohydrolases, the key enzymes in the hydrolysis of crystalline cellulose. In addition, the presented work also describes the characterization of novel, thermostable cellulases.

1.1 Cellulose

Most of the organic carbon on earth is bound up in plant material, and it is estimated that globally between 10^{10} and 10^{11} tonnes of cellulose are synthesized and also hydrolysed every year (Hon, 1994). The cellulose content of plant cell walls varies generally between 35 and 50 %. The crystalline cellulose fibrils in cell walls (of wood) are about 2.5 nm in diameter and are embedded in an amorphous hemicellulose–lignin matrix. Cellulose is found almost exclusively in plant cell walls, although it is also produced by some algae (e.g. *Cladophora* sp.), animals (e.g. tunicates) and a few bacteria e.g. *Acetobacter xylinum* (Atalla and Van der Hart, 1984; Ross *et al.*, 1991). The complex structure of plant cell

walls is the dominant limiting factor for utilization of untreated raw materials of plant origin, but cellulolytic fungi and bacteria living in plant litter have evolved to degrade these materials in order to obtain a supply of energy and carbon. The most recalcitrant component of plant cell walls, lignin, is known to be efficiently depolymerised only by white rot fungi although it is generally believed that lignin depolymerization is necessary to gain access to cellulose and hemicellulose (Cullen and Kersten, 2004).

1.1.1 Crystal structure of cellulose

Cellulose is chemically a simple molecule, as it is a linear monopolysaccharide of anhydroglucose residues connected by β-1,4 linkages. Each residue is tilted by 180° with respect to its neighbours, making anhydrocellobiose the smallest structural subunit (Fig. 1). A cellulose molecule forms hydrogen bonds with itself (intra-molecular H-bonds) and with other cellulose chains (inter-molecular H-bonds), which induces spontaneous crystallisation of the chains to form bundles called microfibrils. The microfibrils associate to form fibrils that finally build up the cellulose fibre (Hon, 1994), which has a structural role in the cell wall, providing strength and contributing to the shape of the cell. A recent investigation by Ding and Himmel (Ding and Himmel, 2006) suggested that the cellulose microfibril in plant cell walls is composed of 36 cellulose chains with a total of six sheets only. Of these 36 chains, only the inner chains are crystalline in nature, whereas the outer chains are non-crystalline.

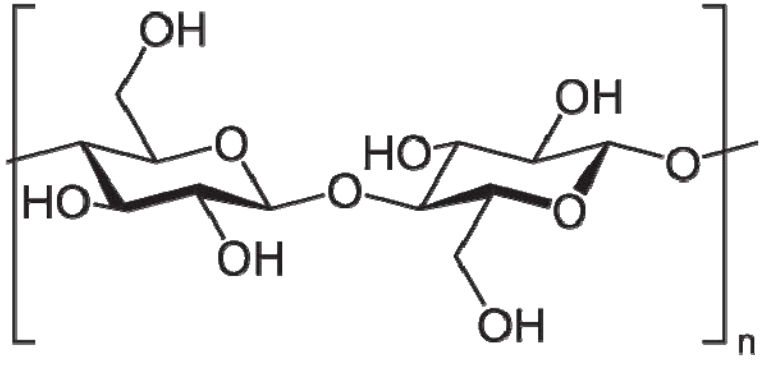

Fig. 1. Molecular structure of cellulose (n = DP, degree of polymerisation).

Pure cellulose can exist in different crystalline forms. Native crystalline cellulose, which is called cellulose I can exist in two different crystalline forms: cellulose I_α and cellulose I_β (Fig. 2). The proportions of these forms vary depending on the source of cellulose, I_α being more prominent in celluloses of algal and bacterial origin and I_β in higher plants (Atalla and Van der Hart, 1984). The fundamental difference between these allomorphs is that I_α is less stable and can be irreversibly converted into I_β by hydrothermal treatment (Sugiyama *et al.*, 1990; Yamamoto *et al.*, 1989). The structures of cellulose I_α and I_β have been determined using synchrotron and neutron diffraction data; cellulose I_α arranges in a triclinic and cellulose I_β in a monoclinic lattice. Within these celluloses, the chains are arranged in a parallel orientation to form layers, which are stabilised by inter- and intra-molecular hydrogen bonds and by van der Waals interactions. The layers are furthermore stacked to form the crystal lattice. The layer stacking has been thought to be due to only van der Waals interactions, but the atomic structures, solved by synchrotron X-ray and neutron diffraction, have shown that weak C-H···O hydrogen bonding also plays a role (Nishiyama *et al.*, 2002; Nishiyama *et al.*, 2003). There are more C-H···O inter-sheet bonds in I_β than in I_α. Differences in the crystalline organization affect the susceptibility to enzymatic hydrolysis and different hydrolytic rates have been observed, I_α being an easier substrate for the Cel7A type cellobiohydrolases studied in this thesis (Hayashi *et al.*, 1997; Igarashi *et al.*, 2006). In addition to the native cellulose I, other crystalline forms also exist (II, III and IV), of which cellulose II is the most thermodynamically stable due to its altered hydrogen bonding network. The other forms of cellulose are produced from cellulose I by different physico-chemical treatments and do not exist in nature (Klemm *et al.*, 2005).

The crystalline structure not only makes cellulose resistant to enzymatic hydrolysis, but also prevents penetration of water and other small molecules. However, the crystalline regions contain many irregularities, such as dislocations and twists of the cellulose chains. As a result of the structural heterogeneity, some parts of the fibres can be partially hydrated by water and some parts are open enough for cellulolytic enzymes to penetrate and gain access to the cellulose chains.

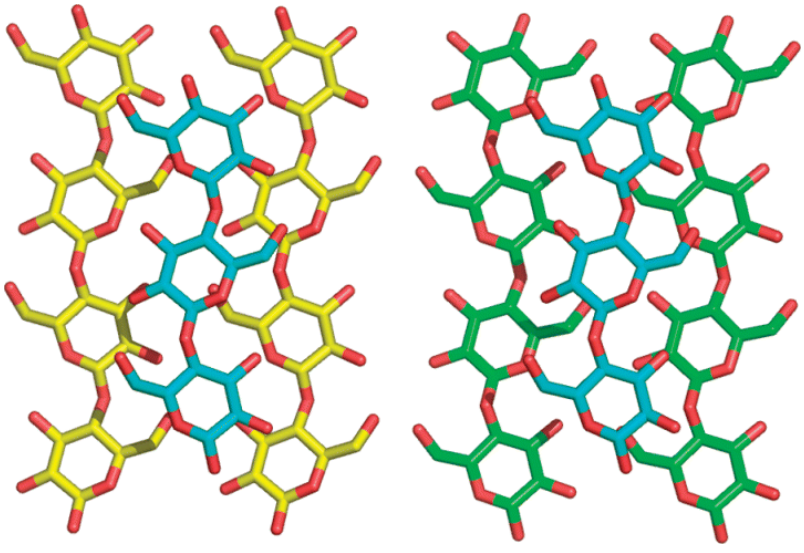

Fig. 2. Views of the hydrophobic surfaces of cellulose I_α (left) and cellulose I_β (right). The cellulose chains in the first layer are superimposed and colored cyan. The chains in the second layer are colored yellow (cellulose I_α) and green (cellulose I_β). The differences between I_α and I_β are very subtle, cellulose I_β having a slightly higher density than that of I_α. The picture has been adopted from (Igarashi et al., 2006).

1.2 Enzymatic degradation of cellulose

Hydrolysis of cellulose in sufficient order of magnitude to support growth is characteristic for microorganisms, which use glucose released from cellulose as carbon source for their growth. However, some animal species, including termites, are also known to produce cellulases, although the contribution of cellulose in their nutrition is still unclear (Watanabe and Tokuda, 2001). Because of the polymeric and crystalline nature of cellulose, it is evident that cellulolytic microbes are not able to take the substrate up into the cells, and therefore the degrading enzymes are either located on the cell surface or are produced extracellulary. Microorganisms have different courses of action to utilize cellulose as their carbon source. Aerobic cellulolytic filamentous fungi can penetrate cellulosic substrates with their hyphae and thereby secrete their cellulase system into the substrate material (Eriksson *et al.*, 1990). Cellulolytic anaerobic bacteria (as well as some anaerobic fungi in the gastrointestinal tracts of ruminant animals) possess an alternative mechanism of cellulose degradation through production of cellulosomes, cellulase complexes anchored to their cell wall.

Cellulases are O-glycoside hydrolases (GH) and they hydrolyse the β-1→4 glycosidic bonds in cellulose. Cellulases have traditionally been classified on the basis of their mode of action into endo- and exoglucanases. Endoglucanases (1,4-β-D-glucan glucanohydrolase; EC 3.2.1.4) can hydrolyse cellulose chains in the middle by making random cuts, and are able to attack only the non-crystalline, amorphous regions (Teeri, 1997). Endoglucanases efficiently reduce the cellulose chain length (DP) and create new chain ends for the action of exoglucanases. Exoglucanases or cellobiohydrolases (1,4-β-D-glucan cellobiohydrolase; EC 3.2.1.91) hydrolyze cellulose chains sequentially, by removing cellobiose units from the reducing or non-reducing end of the cellulose chain in a processive manner (Teeri *et al.*, 1998; Wood and Bhat, 1988). They are usually active on both crystalline and amorphous parts of the cellulose but decrease the DP of cellulose very slowly.

For efficient degradation of crystalline cellulose, usually both endo- and exoglucanases are needed (Henrissat and Bairoch, 1996). β-Glucosidases (EC 3.2.1.21) complete the cellulose hydrolysis by cleaving the resulting cellobiose into glucose. The concerted action of three different types of cellulases as a non-complexed enzyme system is presented in Fig. 3.

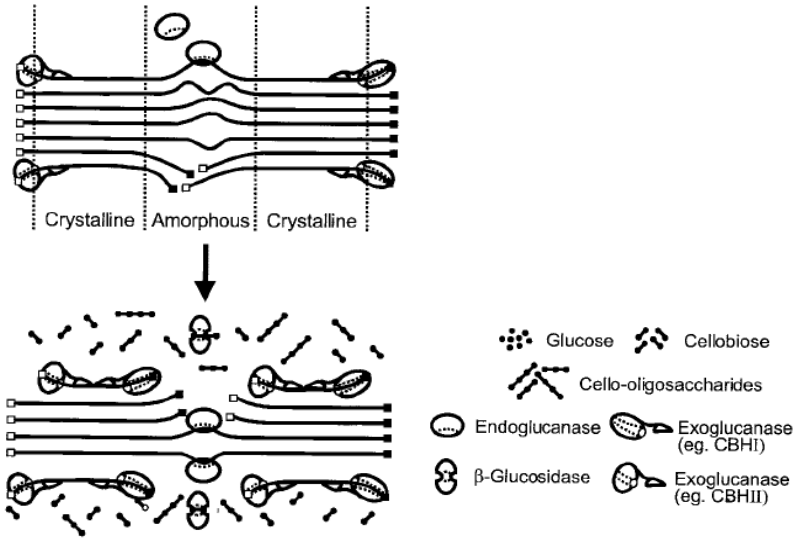

Fig. 3. Schematic representation of the hydrolysis of amorphous and microcrystalline cellulose by non-complexed cellulase systems. The solid squares represent reducing ends and the open squares non-reducing ends. Amorphous and crystalline regions are indicated (Lynd et al., 2002).

Differing from the cellulolytic system presented in Fig. 3, some bacterial cellulolytic systems have been reported to have processive endoglucanases. Moreover, e.g. *Humicola insolens* cellobiohydrolase Cel6A has been reported to be capable of both exo- and endo-actions, (Boisset *et al.*, 2000), thus the division to endo- and exo-acting enzymes may not be fully satisfactory.

1.2.1 Free/non-complexed and complexed cellulase systems

Most aerobic cellulolytic microbes secrete sets of individual cellulases. The most studied free cellulase systems are those of the aerobic fungal species *Trichoderma reesei, Humicola insolens* and *Phanerochaete chrysosporium.* They all produce a complete set of cellulases able to hydrolyse native cellulose effectively. *T. reesei* and *H. insolens* both produce two cellobiohydrolases (Cel7A and Cel6A), while *P. chrysosporium* produces seven cellobiohydrolases, Cel6A and Cel7 isoenzymes from A to F (Kubicek, 1992; Munoz *et al.*, 2001; Schülein, 1997). In addition to the exoglucanases, they all produce several endoglucanases (belonging to GH families 5, 7, 9, 12 and 45) (Okada et al., 1998; Penttilä et al., 1986; Saloheimo et al., 1994; Saloheimo et al., 1997; Saloheimo et al., 1988; Schülein, 1997; Wymelenberg et al., 2002) and β-glucosidases [Cel1(intracellular) and Cel3 (extracellular)] (Foreman *et al.*, 2003; Saloheimo *et al.*, 2002a; Takashima *et al.*, 1999; Tsukada *et al.*, 2006). The complete genomes of *T. reesei* and *P. chrysosporium* have been published and the *P. chrysosporium* data provides information of a number of putative cellulases besides the above mentioned enzymes. *P. chrysosporium* genome contains the genetic information to encode at least 40 putative endoglucanases and seven cellobiohydrolases, while *T. reesei* genome unexpectedy contains only ten genes coding endoglucanases and cellobiohydrolases (Martinez et al., 2004; Martinez et al., 2008).

In addition to the above mentioned mesophilic fungi, some thermophilic fungi are also capable of decomposing cellulose. The cellulolytic system of *Talaromyces emersonii*, a moderately thermophilic fungus, has hitherto been reported to include two Cel7 exoglucanases, one Cel6 exoglucanase (Tuohy *et al.*, 2002), endoglucanases (Moloney *et al.*, 1985), and one extracellular and one intracellular β-glucosidase (McHale and Coughlan, 1981). The sequence and detailed biochemical characterization of these *T. emersonii* cellobiohydrolases have shown comparable properties with the cellobiohydrolases of *T. reesei* and *P. chrysosporium* (Tuohy *et al.*, 2002). The cellulase enzymes of another

thermophilic fungus, *Melanocarpus albomyces*, include at least two endoglucanases (Cel45A and Cel7A) and one cellobiohydrolase (Cel7B). These enzymes have been cloned and characterized for their pH and temperature behaviour (Miettinen-Oinonen *et al.*, 2004).

One of the most studied species of cellulolytic aerobic bacteria, *Cellulomonas fimi*, produces multible endoglucanases and exoglucanases (Meinke *et al.*, 1993; Meinke *et al.*, 1994; Shen *et al.*, 1995). The individual cellulases of *Cellulomonas* are similar to those of the cellulase systems of aerobic fungi, although clustering of the enzymes through scaffolding to form cellulosome-like protuberant structures has been observed on *Cellulomonas* cells grown with cellulose and cellobiose as carbon sources (Lamed *et al.*, 1987).

Most cellulolytic anaerobic bacteria and fungi possess a cellulosome (Lynd *et al.*, 2002). Cellulosomes are stable enzyme complexes, which are firmly bound on the bacterial cell wall but are however flexible enough to bind at the same time onto cellulose. Cellulosomes can be described as macromolecular machines designed to hydrolyse insoluble cellulose substrate. The most studied anaerobic cellulolytic microorganism is the thermophilic bacterium *Clostridium thermocellum*, which produces an extracellular multi-enzyme complex, reviewed by (Bayer et al., 1998; Bayer et al., 2004; Schwarz, 2001). The size of the cellulosomes varies between 2000 kDa and 6500 kDa and they are composed of up to 50 components detected as protein bands on SDS-PAGE gel, depending on the strain and the growth conditions. The largest subunit of the cellulosome is a non-catalytic scaffoldin protein CipA, which comprises 9 cohesin domains, a single cellulose binding module CBMIII, a dockerin domain and hydrophilic modules designed to hold the complex together. The scaffoldin is anchored to the cell wall via type II cohesin domains. Altogether 22 catalytic modules of *C. thermocellum*, of which at least 13 exhibit cellulose activity, have dockerin moieties that can associate tightly with the cohesins of the CipA protein to form the cellulosome. It is generally believed that the cellulosome composition can vary depending on the strain and the growth conditions.

1.2.2 Synergy

Individual cellulases are not able to hydrolyse crystalline cellulose efficiently, and the cooperative action of different types of cellulases is needed. Synergism refers to the ability of a mixture of cellulases to have higher activity than the sum of the activities of the individual enzymes. Synergism occurs when the

different cellulases attack the substrates at different sites, thus creating or revealing new sites for other enzymes to work on. The two main types of cellulase synergies are endo-exo synergy and exo-exo synergy. Endo-exo synergy is explained by the endo-acting cellulase forming new (reducing and non-reducing) chain ends for cellobiohydrolases. Endo-exo synergy has been reported e.g. for the four major cellulases of *T. reesei*, Cel6A (former CBHII), Cel7A (former CBHI), Cel7B (former EGI) and Cel5A (former EGII) has been studied, Cel7B and Cel5A act synergistically with both Cel6A and Cel7A (Medve et al., 1994; Medve et al., 1998; Nidetzky et al., 1994a). Processive endocellulases, in contrast to true endoglucanases, can exhibit synergism with both endo- and exo-acting cellulases, as detected for the *Thermobifida fusca* processive endocellulase Cel9A (Wilson, 2004). Exocellulases have also been shown to act synergistically with other exoglucanases (Medve *et al.*, 1994). The explanation for this type of synergy is not as simple on a molecular level as endo-exo synergy and two putative causes have been discussed: 1) exocellulase synergism is due to the different chain end specificities and directionalities (Hoshino *et al.*, 1997; Irwin *et al.*, 1993; Nidetzky *et al.*, 1994a), 2) exocellulase synergism is due to the inherent endo -character of some cellobiohydrolases (Boisset *et al.*, 2000).

1.2.3 Cellulases are modular proteins

Most cellulases show a modular architecture including one or more catalytic modules and one or more modules involved in substrate binding (CBMs, carbohydrate-binding modules) or multienzyme complex formation (Davies and Henrissat, 1995). In addition to these modules, at least bacterial cellulases may also have additional modules e.g. homologous to fibronectin typeIII (FnIII) or immunoglobulin (Bayer *et al.*, 2006). The FnIII-like module of *Clostridium thermocellum* has been shown to promote cellulose hydrolysis by loosening the surface structure of the substrate (Kataeva *et al.*, 2002). The distinct modules are often connected via linker peptides of varying length. The modular structure of fungal cellulases was discovered in the 1980s by limited proteolysis studies, which revealed the existence of separate binding and catalytic modules connected by a linker peptide (Tomme *et al.*, 1988b; van Tilbeurgh *et al.*, 1986). The flexibility and heterogenic O-glycosylation of the linker regions is presumed to be the major obstacle to protein crystallization of intact fungal cellobiohydrolases containing both the catalytic module and the cellulose binding module, and

therefore there are no crystal structures available presenting the two modules connected by a linker. The linker regions vary considerably in length, and in many cases little or no sequence identity is found between the linkers from different organisms, although they are often rich in proline and serine or threonine residues (Gilkes *et al.*, 1991). The cooperativity between the catalytic and cellulose-binding module can be accounted for by intra-molecular synergy. The *T. reesei* Cel7A catalytic module has been shown to act in concert with CBM, and the spatial separation between the modules appears to be important for the mechanistic action of the enzyme (Linder *et al.*, 1996; Palonen *et al.*, 1999; Srisodsuk *et al.*, 1993).

1.2.4 Families of cellulase catalytic modules

The catalytic modules of different glycoside hydrolases are classified into over 100 families based on amino acid sequence similarities (Henrissat and Bairoch, 1996). The CAZy (Carbohydrate Active enzymes) database (http://www.cazy.org/) collates different cellulolytic enzymes currently into 14 families (families 5–10, 12, 26, 44, 45, 48, 51, 61, 74), listed in Table 1. There is usually a direct relationship between the amino acid sequence and protein fold, and therefore the classification of glycoside hydrolases provides information about the protein fold and the catalytic mechanism. The addition of information on the 3-dimensional structure of many proteins in the classification has confirmed that each family contains one basic protein fold (Henrissat *et al.*, 1995). To date, all the GH families to which cellulolytic enzymes have been allocated also contain enzymes for which 3D structural information is available.

The comparison of the 3D structures of glycoside hydrolases has revealed three different types of structural arrangements for the glycoside hydrolase active site: 1) a tunnel, which is suitable for processive exo-hydrolysis, 2) a cleft suitable for endo-attack, 3) a pocket (Davies and Henrissat, 1995). Figure 4 presents examples of these three different shapes of catalytic sites. A tunnel-shaped active site was first revealed in *T. reesei* Cel6A, an α/β-protein with a barrel topology, in which the active site is located in an enclosed tunnel formed by two surface loops (Koivula *et al.*, 2002; Rouvinen *et al.*, 1990; Zou *et al.*, 1999) (Fig. 4A). The exo-action of this enzyme was explained by the tunnel shaped active site. However, apparently the two loops of Cel6 cellobiohydrolase, covering the active site, can occasionally open to allow an endo-type of initial action (Boisset *et al.*, 2000). The 3D structures of endo-active enzymes from the

same family have shown that the same fold can also form a more open, cleft-like active site. In this case the loops forming the tunnel are shorter than in *Tr* Cel6A or are turned away from the active site (Fig. 4B). β-Glucosidase from the GH-1 family serves as an example of a glycoside hydrolase having a pocket-shaped active site (Fig. 4C).

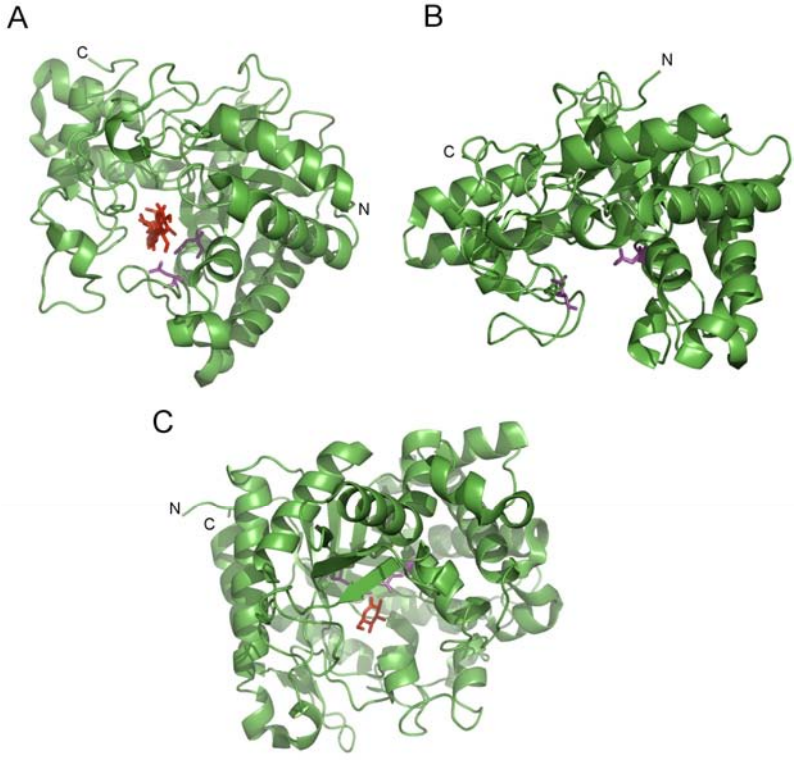

Fig. 4. Three-dimensional structures of A) *T. reesei* Cel6A (Rouvinen et al., 1990), B) *H. insolens* Cel6B (Davies et al., 2000) and C) Phanerochaete chrysosporium Bgl1 (Nijikken et al., 2007) demonstrating the three different catalytic site shapes of glycoside hydrolases described by (Davies and Henrissat, 1995). The picture was created with the program PyMol (DeLano Scientific, San Carlos, CA, USA).

In the GH-7 family a total of seven crystal structures have been solved, including four cellobiohydrolases: *T. reesei* Cel7A (Divne *et al.*, 1994), *T. emersonii* Cel7A (Grassick *et al.*, 2004), *P. chrysosporium* Cel7D (Munoz *et al.*, 2001) and *M. albomyces* Cel7B (Parkkinen *et al.*, 2008) and three endoglucanases: *H. insolens* Cel7B (Davies *et al.*, 1997), *T. reesei* Cel7B (Kleywegt *et al.*, 1997) and *Fusarium oxysporum* Cel7B (Sulzenbacher *et al.*, 1996). The overall fold of

these enzymes is conserved; however, the active site of cellobiohydrolases is tunnel-shaped whereas endoglucanases have a more open, cleft-shaped active site, similar to those observed in GH family 6. The folds of both exo- and endoglucosidases of GH family 7 consist of an antiparallel β-sandwich, but the six extended surface loops, which form the active site tunnel in exoglucanases, are shorter in endoglucanases. The cellulose binding tunnel of *Tr* Cel7A contains 11 subsites (-7 → +3) which can accommodate up to 10 glucose units of the cellulose chain (Divne *et al.*, 1998). The catalytic site is between subsites -1 and +1, and the sites +1→+3 are so-called product sites (Fig. 5).

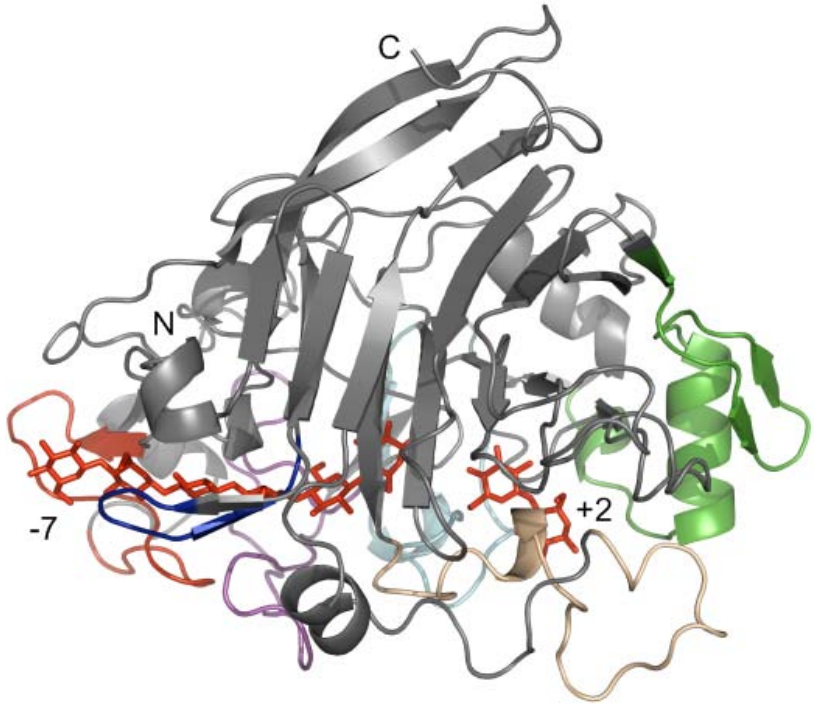

Fig. 5. 3D structure of the *T. reesei* Cel7A in complex with cello-oligosaccharides (PDB code 7cel) showing the tunnel forming loops: Loop 1 (T41-T59) in red, Loop 2 (G98-G105) in blue, Loop 3 (R178-G205) in violet, Loop 4 (T231-T254) in light blue, Loop 5 (T310-K346) in green and Loop 6 (D369-R394) in beige, and a cellulose chain bound in the active site tunnel and spanning the subsites -7 to +2.

Table 1. The glycoside hydrolase families containing published cellulase gene sequences in CAZy (http://www.cazy.org/).

Family	Mechanism	Type	EC No.
5	Retaining	Endoglucanase	3.2.1.4
6	Inverting	Endoglucanase	3.2.1.4
6	Inverting	Cellobiohydrolase	3.2.1.91
7	Retaining	Endoglucanase	3.2.1.4
7	Retaining	Cellobiohydrolase	3.2.1.91
8	Inverting	Endoglucanase	3.2.1.4
9	Inverting	Endoglucanase	3.2.1.4
10	Retaining	Endoglucanase	-
12	Retaining	Endoglucanase	3.2.1.4
26	Retaining	Endoglucanase	3.2.1.4
44	Inverting	Endoglucanase	3.2.1.4
45	Inverting	Endoglucanase	3.2.1.4
48	Inverting	Cellobiohydrolase	3.2.1.91
51	Retaining	Endoglucanase	3.2.1.4
74	Inverting	Endoglucanase	3.2.1.4

1.2.5 Families of carbohydrate-binding modules

The CBMs, similarly to the catalytic modules of the glycoside hydrolases, can be classified into different families (http://www.cazy.org/). To date this classification includes 57 CBM families. The classification of CBMs has broadened from cellulose binding domains to carbohydrate binding modules, since additional modules in carbohydrate-active enzymes which bind carbohydrates other than cellulose are continuously found. The CBM affects the binding of cellulases to the cellulose surface. Apparently, CBM aligns the catalytic module close to the polymeric substrate, prolonging the association with the substrate and thereby increasing the effective enzyme concentration (Igarashi *et al.*, 2009). The presence of CBM is especially important for the activity of exoglucanases and experimental evidence has shown that removal of the binding module dramatically decreases the activity on crystalline substrates (see (Gilkes *et al.*, 1991; Shoseyov *et al.*, 2006) for a review). In addition, there is evidence that some CBMs are also capable of disrupting the crystalline structure of cellulose (Boraston *et al.*, 2004).

CBMs can be grouped into seven different fold families, but here the fold is not predictive of function because of the diversity in the specific amino acids and binding site topographies. The binding specificity is determined by binding site topography, and interaction with a ligand and ligand recognition is invariably due to aromatic amino acids that form the CBM binding sites. Binding sites can be planar (as in CBM family 1), twisted or sandwich-shaped (see (Boraston *et al.*, 2004) for a review). The CBMs of fungal cellulases are invariably small modules, approximately 40 amino acids in size, stabilized by two or three disulphide bridges and belong to the CBM-1 family. Cellulose-binding function has been demonstrated in many cases, and appears to be mediated by three aromatic residues separated by about 10.4 Å forming the flat surface (Linder *et al.*, 1995; Reinikainen *et al.*, 1992). The 3D structure of *T. reesei* Cel7A CBM is presented in Fig. 6.

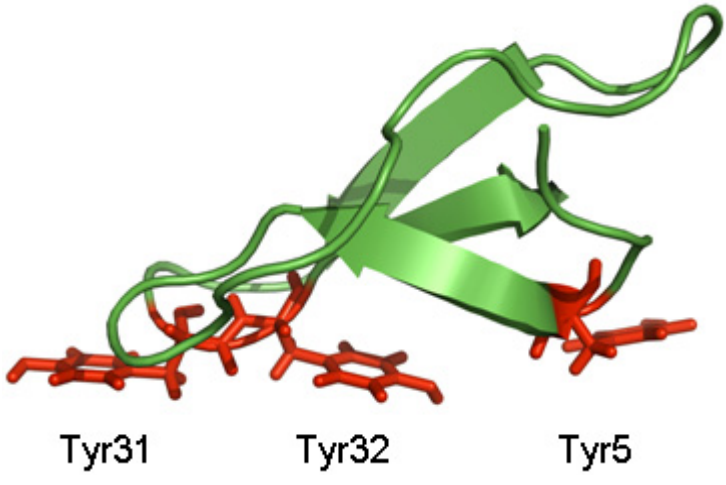

Fig. 6. 3D structure of *T. reesei* Cel7A CBM (PDB code 1cbh) determined by solution NMR (Kraulis et al., 1989). The aromatic amino acids in the binding face are shown in red. The picture was created with PyMol.

1.2.6 Catalytic mechanisms

Enzymatic hydrolysis of glycosidic bonds proceeds through general acid catalysis and it occurs (by definition) between the carbohydrate units bound at the subsites -1 and +1. This is a stereoselective process in which the configuration about the anomeric centre (at the C1 carbon) can either be inverted

or retained. Different mechanisms are required for these two different stereochemical courses (Hehre, 2000; Koshland, 1953). The general scheme for the reaction mechanisms is shown in Fig. 7. Both reaction mechanisms involve oxocarbonium ion-like transition states. In addition, ring distortion upon substrate binding at the subsite preceding the scissile bond (-1) has been proposed as an element for both reaction types of enzymes (Harris *et al.*, 1993; Koivula *et al.*, 1996; Koivula *et al.*, 2002; Zou *et al.*, 1999). One important difference between the two mechanisms is that most retaining enzymes can catalyze both transglycosylation and hydrolysis, whereas no known inverting enzyme catalyzes transglycosylation (Withers, 2001). The catalytic mechanism appears to be conserved in each glycoside hydrolase family, and therefore both the stereochemistry of the hydrolysis and the fold are conserved within the GH family in question (Henrissat *et al.*, 1998).

Both mechanisms are supposed to require that two carboxylic acid residues are suitably positioned on either side of the bond to be cleaved. Inverting enzymes use a single-displacement mechanism and retaining enzymes a double-displacement mechanism. In the inverting glycosidases the two catalytic carboxylic acid usually lie further apart from each other than in the retaining hydrolases (McCarter and Withers, 1994). In the inverting mechanism the catalytic acid residue donates a proton to the anomeric C1 carbon while the catalytic base residue removes a proton from a water molecule, thus facilitating its nucleophilicic attack. There is an increasing amount of evidence demonstrating that not all glycosyl hydrolases follow the classical catalytic mechanisms. Concerning the inverting enzymes, there are several reports on the lack of a typical catalytic base (as reviewed in Vuong and Wilson, 2010). In the case of e.g. *T. reesei* cellobiohydrolase Cel6A only one catalytic carboxylic acid has been found and the reaction mechamism has been shown to include conformational changes in the active site tunnel (Koivula *et al.*, 2002).

Catalysis by retaining glycosidases has been suggested to proceed via a double-displacement mechanism in which a covalent glycosyl–enzyme intermediate is formed (Withers, 2001). The covalent glycosyl-enzyme intermediate has been shown to be formed in some cases, however, it may not always be needed (Parkkinen *et al.*, 2008; Sinnott, 1990). The active site again contains a pair of carboxylic acids, but in this case they are approximately 5.5Å apart.

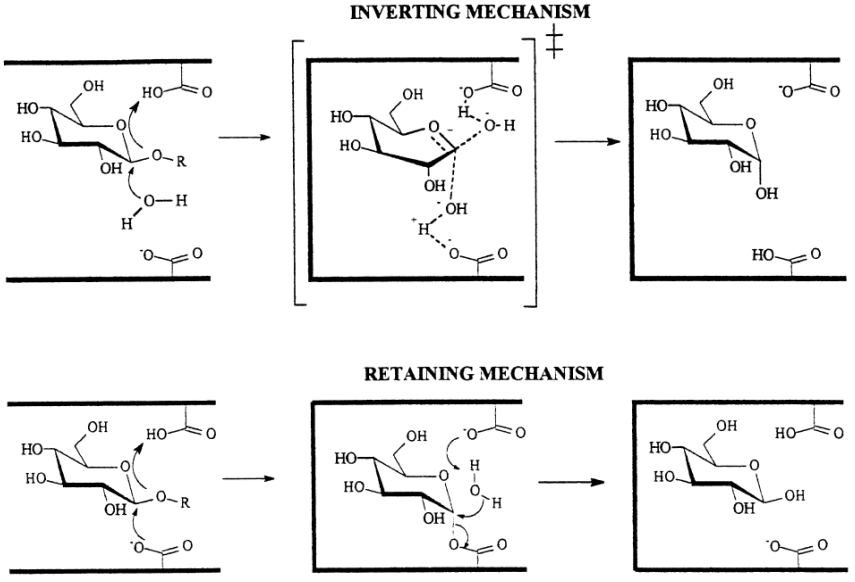

Fig. 7. General scheme for the reaction mechanisms of inverting and retaining glycoside hydrolases. Both pathways pass through transition states with a strong oxocarbonium-ion-like character. From (Withers, 2001).

1.3 Biochemical characterization, enzymatic assays

In order to understand the function of cellulase systems comprised of multiple components, it is important to understand the distinct modes of action (activities) of the individual enzymes. The activity of the individual enzymes, as well as of the whole cellulase system, is affected by substrate properties such as crystallinity, available surface area and degree of polymerization. These substrate properties change during the hydrolysis, e.g. the number of chain ends increases because of the action of endoglucanases (Kleman-Leyer *et al.*, 1996; Srisodsuk *et al.*, 1998). In addition the accessibility of cellulose can vary during the hydrolysis because of the overall substrate conversion (Boisset *et al.*, 2000; Saloheimo *et al.*, 2002b). These changes in the substrate characteristics during hydrolysis cause changes in the hydrolysis rates during the time course of the overall reaction.

Enzymological characterization of individual cellulases is complicated due to the complex physical characteristics of cellulase substrates and the complexity of the action of different cellulases. Therefore, several assay procedures have been developed for the measurement of cellulase activity. Significant differences

in substrate characteristics and in the assay procedures make the comparison of results from different cellulase assays difficult. In addition, most cellulase assays on insoluble substrates are non-linear both with time and with the amount of enzyme, and although activity does increase when the substrate concentration is increased, activity on insoluble cellulose does not show Michaelis-Menten type kinetics (Wilson, 2008).

1.3.1 Substrates for cellulase activity assays

1.3.1.1 Soluble substrates

Soluble substrates include substituted, long chain cellulose derivatives and low DP cello-oligosaccharides and their derivatives. These are often used to determine the cellulolytic activities of individual enzyme components. Cellulose polymers can be solubilised for example by carboxymethyl substitution. The solubility of carboxymethyl cellulose (CMC) depends on the degree of substitution; it starts to be water soluble when DS > 0.3–0.7, i.e. when the molecule contains an average of 3 to 7 carboxymethyl groups per 10 anhydroglucose units. Endoglucanases hydrolyse such substrates by cleaving the bond between substituted groups. The activity can be measured by detecting e.g. reducing sugars or as a decrease in viscosity (Hulme, 1988). Cello-oligosaccharides with a degree of polymerisation up to six glucose units are soluble in water, and they can be used as model substrates for cellulases. The hydrolysis products can be detected with HPLC (Harjunpää *et al.*, 1996; Nidetzky *et al.*, 1994b).

For simplified detection of hydrolysis products, the oligosaccharides can be coupled to chromophoric, fluorogenic or radioactive labels (Claeyssens *et al.*, 1989; Freer, 1993; van Tilbeurgh *et al.*, 1988; Zhang and Lynd, 2006). Commonly used labelled oligosaccharides include 4’-methylumbelliferyl-, 4’-nitrophenyl- or 2-Chloro-4-nitrophenyl-β-D-glycosides derived from D-glucose, cellobiose, higher cello-oligosaccharides, or lactose. Cello-oligosaccharides are useful in determination of initial cellulase kinetics or inhibition constants (Ståhlberg *et al.*, 1996; Tuohy *et al.*, 2002; van Tilbeurgh and Claeyssens, 1985). The data obtained from the model substrates do not correlate directly with crystalline cellulose degradation, in which the efficiency of an enzyme depends on the ability to access single glucan chains on a crystalline structure.

1.3.1.2 Insoluble substrates

Insoluble substrates for cellulase activity measurements are prepared from celluloses from different sources. Cotton, filter paper (Whatman no. 1), bacterial microcrystalline cellulose (BMCC), Avicel and amorphous cellulose are almost pure celluloses having different DP, crystallinity index and accessible surface area, whereas pre-treated lignocellulosic substrates also contain other polysaccharides, lignin and impurities. Avicel is prepared by hydrolysing softwood pulp with dilute hydrochloric acid to remove the amorphous fraction followed by spray drying. Despite the hydrolysis treatment Avicel contains a substantial amount (30–50%) (Zhang and Lynd, 2004) of amorphous cellulose. Amorphous cellulose can be prepared for example from Avicel by mechanical or chemical methods. The properties of amorphous cellulose vary greatly depending on the preparation method e.g. reaction temperature and time, and the hydrolysis results conducted with different batches made even in the same laboratory can not always be reliably compared (Zhang et al., 2006).

The detection of cellulase activity on an insoluble substrate can be performed by following the DP decrease, or more often by measuring the released products, reducing sugars. Reducing sugar assays including the 3,5-dinitrosalicylic acid (DNS) method (Ghose, 1987; Miller, 1959), the Nelson-Somogyi method (Nelson, 1944), the 2,2´-bicinchroninate (BCA) method (Johnston *et al.*, 2007), the para-hydroxybenzoic acid hydrazide (PAHBAH) method (Lever, 1972) and the ferricyanide method (Kidby and Davidson, 1973), of which the DNS method is the most extensively used. However, the DNS method is rather insensitive and subject to interference by many compounds, e.g. metals and phosphates in cultivation media (Dighe *et al.*, 1985; Forouchi and Gunn, 1983). It should also be noted that different reducing sugars generally yield different colour intensities; thus, it is necessary to calibrate for each individual sugar. PAHBAH and BCA methods are reported to be more sensitive and reliable. If a cellulase assay is performed with a complete cellulase system or a mixture also including a β-glucosidase, the hydrolysis can also be detected by measuring released glucose in a coupled enzymatic assay applying e.g. glucose oxidase and peroxidase (Kunst *et al.*, 1984).

1.4 Industrial applications of cellulases

Carbohydrases such as cellulases and amylases are currently the second most important group of industrial enzymes after proteases based on sales (Kirk *et al.*, 2002). Cellulases have proved to be commercially useful in the textile industry, substituting conventional stonewash methods (biostoning). They also have applications in the detergent industry for reducing fuzz and pilling of fabrics, and as general cleaning agents for cotton garments (Bhat, 2000; Olson and Stanley, 1990). In addition, they are used in animal feeds for improving their nutritional quality and digestibility, and in the processing of fruit juices (Bhat, 2000). They have great potential in the paper and pulp industries, and have been used to lower the energy consumption in mechanical pulp production (Pere *et al.*, 2000).

Cellulase research has recently been focused particularly on the industrial conversion of cellulose-containing biomass (lignocellulose) to fermentable sugars, which can be converted to ethanol to serve as an alternative to fossil fuels (Duff and Murray, 1996). Production of ethanol from annual crops (e.g. starch) is already a well established technology. Starch- and sugar cane-based ethanol is often referred to as first-generation biofuel (Yuan et al., 2008). However, there is a need to have an alternative, non-food raw material to avoid the conflict between land use for food and for energy production. Cellulose-based ethanol is a promising alternative to reduce oil dependence as well as greenhouse gas emissions, because the plant biomass raw material is renewable, inexpensive and abundantly available. Biomass sources such as agricultural waste (straw), forest wood, side-products of the forest industry and energy crops are the main potential sources. The biomass supply in the EU has been estimated to be high enough to contribute significantly to the total energy supply in Europe, and the future of bioethanol production from lignocellulose mainly depends on agricultural policy (Ericsson and Nilsson, 2006; Gnansounou, 2010). Currently, the processes using lignocellulosic feedstock for bioethanol production are still in the development phase, due to the challenges in hydrolysis of the recalcitrant lignocellulosic substrates. The natural cellulose decomposers, cellulolytic bacteria and fungi, are able to utilize lignocellulose as a carbon source by hydrolysing the material with secreted cellulolytic enzymes. This hydrolysis process is slow, because lignin and the crystalline nature of the substrate restrict access of the enzymes to the polysaccharide. Rapid and efficient hydrolysis of lignocellulose for industrial purposes requires pretreatment to open the resistant structure of the feedstock and to make it more susceptible to enzyme attack.

Pretreatment processes typically involve elevated temperature and pressure combined with acid or base catalysis (Kim and Lee, 2007; Mais *et al.*, 2002; Palonen *et al.*, 2004). The major obstacles to cost-effective production of ethanol from cellulose are the high enzyme costs due to the high amount of enzymes needed in the process. One option to tackle this problem is to improve the currently used enzyme mixtures by improving the individual enzymes by protein engineering or enzyme discovery, or by optimising the enzyme mixtures e.g. by adding novel enzymes with additional activities. The currently used lignocellulosic enzymes for bioethanol production are produced in industrially relevant production strains of filamentous fungi, such as *Trichoderma reesei*. Bioethanol process concepts can generally be described as separate hydrolysis and fermentation (SHF) or as simultaneous saccharification and fermentation (SSF) (Viikari *et al.*, 2007). Due to the instability of *T. reesei* enzymes at high temperatures, the hydrolysis process of SHF can be performed at 40–50°C followed by a separate fermentation step at a lower temperature. By using thermostable enzymes the SHF process could be more efficient due to more efficient hydrolysis at higher temperatures. Although the enzymatic hydrolysis in SSF is performed in lower temperature simultaneously with the fermentation, thermostable enzymes could be advantageous, as they are generally robust, tolerating various harsh process conditions.

1.5 Thermostability of industrial enzymes

Thermostable enzymes are suitable for biotechnical applications and are likely to be economically beneficial in industrial processes. Furthermore, they are interesting from a fundamental point of view when studying the structural basis of thermostability. Thermostability is a beneficial property for enzyme applications, since running an industrial process at high temperature has several advantages, such as higher solubility of reactants, reduced risk of microbial contamination and potentially higher reaction rates. The general robustness and better stability of thermophilic enzymes compared to mesophilic ones also increases the recyclability of enzymes in the industrial process.

A generalisation of the Arrhenius law postulates that many reaction rates of chemical reactions are approximately doubled when temperature is increased by 10°C; thus the reaction rates of hyperthermophilic enzymes would be expected to be 50–100 times higher than those of mesophilic enzymes (Vieille and Zeikus, 2001). However, the reaction rates observed are often similar to those of

mesophilic enzymes measured at their temperature optima. This means that thermostable enzymes often have low activity at medium temperatures (van den Burg and Eijsink, 2002). It has been suggested that these lower catalytic rates could be due to a significantly increased rigidity of the thermophilic enzymes compared to mesophilic enzymes at mesophilic temperatures (Wrba *et al*., 1990; Jaenicke *et al*, 1996). However, there are examples of significant improvements in stability by protein engineering, without compromising the enzymatic activity in low temperatures (Mansfeld *et al.*, 1997; van den Burg *et al.*, 1998; Williams *et al.*, 1999).

1.5.1 Thermostable enzymes in nature

Microorganims are classified into psychrophiles, mesophiles, thermophiles and hyperthermophiles, depending on their optimal growth temperature. Thermophilic organisms (eubacteria and archea) offer a potential source of thermostable enzymes for industrial applications. Only a few eukaryotic organisms are able to grow at temperatures of 45°C–55°C or above (Maheshwari *et al.*, 2000) and thermophily among fungi is not as common as in eubacteria or archea, some of which are able to grow at temperatures between 80°C and 110°C (Vieille and Zeikus, 2001).

1.5.2 Structural reasons for higher thermostability

Although thermostability is one of the most studied and engineered protein properties, no generally applicable rules have been established so far, and it seems unlikely that a single universal stabilization mechanism is responsible for thermostability. Numerous studies on thermophilic properties of proteins have indicated that some stabilization mechanisms vary depending on the specific enzyme (reviewed in (Li *et al.*, 2005; Sterner and Liebl, 2001; Vieille and Zeikus, 2001). The differences in the primary and the three dimensional structures of thermophilic and mesophilic enzyme homologues are subtle, and mainly determined by their phylogenetic differences. Increased thermostability is thought to result from many small cumulative changes in the protein structure. The catalytic mechanisms can also be similar (Bauer and Kelly, 1998; Vieille *et al.*, 1995), and the main difference between hyperthermophilic and mesophilic variants would appear to be the temperature range over which they are stable. The potentially stabilising features, detected from comparisons of 3D structures

of thermophilic and mesophilic proteins are difficult to distinguish from neutral mutations accumulated during evolution (Jaenicke and Bohm, 1998), and only wide comparisons of amino acid sequences and three dimensional structures are expected to provide statistically relevant results (Sterner and Liebl, 2001). The factors which are known to affect thermostability are electrostatic and hydrophobic interactions, hydrogen- and disulphide bonds, overall rigidity and compactness, stable α-helices and loops, glycosylation and metal binding. These aspects are dealt in more details below.

Electrostatic interactions

The most significant sequence and structural differences detected in comparison studies (Haney et al., 1999; Kumar and Nussinov, 2001; Szilagyi and Zavodszky, 2000) were decrease in the number of uncharged polar residues (Gln, Asn, Thr, Ser) and increase in the number of charged amino acid residues (Asp, Glu, Arg, Lys) in thermophilic proteins, meaning that more thermostable proteins often have more electrostatic interactions than their less thermostable homologues. Typical charge-charge interactions between oppositely charged residues are known as salt bridges. The increase of charged residues indicates that there are more salt bridges in thermostable than in mesophilic proteins (Das and Gerstein, 2000; Karshikoff and Ladenstein, 2001).

Hydrogen bonds

The number of hydrogen bonds can be correlated with the thermostability of proteins, as shown for example by Vogt (Vogt *et al.*, 1997) in a theoretical examination of several protein families with different thermal stabilities. However, because the H-bond identification is dependent on high resolution protein structure, no explicit and statistically relevant estimate of the role of H-bonds is available (Vieille and Zeikus, 2001). Strong, low-barrier type of hydrogen bonds have been shown to contribute to improved stability of *T. reesei* cellobiohydrolase Cel6A at acidic pH (Wohlfahrt *et al.*, 2003).

Hydrophobic bonds

Hydrophobic forces facilitate protein folding and have a major impact on stability (Goodenough and Jenkins, 1991). Hydrophobicity is the ratio of buried non-polar surface area to the total non-polar surface area of the molecule.

Replacement of amino acids with more hydrophobic residues in a hydrophobic environment has been used as a site-directed mutagenesis strategy for example for chloramphenicol acetyltransferase (Chirakkal *et al.*, 2001), resulting in improved thermostability. Thermostabilization of a maltogenic amylase by directed evolution has also been reported to be due to enhanced hydrophobic interactions (Kim *et al.*, 2003).

Disulphide bonds

The stabilizing effect of disulphide bridges is due to decreased conformational entropy. Genetically engineered additional disulphide bridges have been reported to increase protein thermostability by 5–30°C (Pace, 1990). Examples of successful thermostabilization by adding disulphide-bridges were reported for e.g. bacteriophage T4 lysozyme (Matsumura *et al.*, 1989), neutral protease (Mansfeld *et al.*, 1997) and xylanase (Xiong *et al.*, 2004).

Rigidity

Thermostable enzymes are active at high temperatures and they need to be more rigid than mesophilic proteins. The rigidity helps them to maintain the catalytically active conformation and prevents unfolding. Rigidity can be demonstrated by low hydrogen-deuterium exchange rates and reduced thermal unfolding. Rigidity can also make proteins generally more resistant to most common protein denaturants (pH, guanidine-hydrochloride and organic solvents) and decreases their susceptibility to proteolysis (Kanaya and Itaya, 1992; Veronese *et al.*, 1984; Wrba *et al.*, 1990).

Compactness

Enhanced compactness of the protein fold is proposed to be one of the general determinants of a stable structure. The compactness of a protein has been defined as the ratio of accessible surface area to the surface area of a sphere with the same volume as that of the protein. In some studies e.g. (Knapp *et al.*, 1997; Russell *et al.*, 1997), thermostability has been associated with compactness, achieved by the shortening of some loops, an increase in the number of atoms isolated from solvent, optimized packing of side chains in the interior, and absence of cavities. However, statistical examinations of the packing efficiency of thermophilic compared to mesophilic proteins have indicated no such

tendency (Karshikoff and Ladenstein, 1998; Kumar *et al.*, 2000), and compactness values calculated for thermophilic and mesophilic proteins have been very similar.

α-Helix stabilization

Introduction of residues with a high α-helix propensity has been confirmed experimentally to affect protein stability. Alanine residues in α-helices have been shown to stabilize the protein structure (Menendez-Arias and Argos, 1989; Zhang et al., 1995). In addition, possible stabilization of α-helices by replacement of lysine residues in an α-helix by a glutamic acid residue has been proposed (Davies et al., 1993).

Stabilization of loops

Loops are often the regions in protein structures which have the largest thermal factors (B-factors), indicating that they are mobile and likely to unfold first (Daggett and Levitt, 1992; Lazaridis *et al.*, 1997). Proline substitutions (Xxx to Pro) have been used in flexible loop areas to improve thermostability. Proline has the lowest conformational entropy, since it is able to adopt only few configurations. At the same time it also restricts the conformational freedom of the backbone of polypeptide chains and therefore proline substitutions are expected to contribute to protein stability by reducing the entropy of the unfolded states (Dill, 1990; Matthews *et al.*, 1987; Schimmel and Flory, 1968).

Glycosylation

Post-translational modifications like glycosylation and phosporylation are known to influence the protein properties. Glycoporteins are often more stable than their unglycosylated analogues (Imperiali and O'Connor, 1999; Wang et al., 1996). The stabilising effect of glycans has been shown to be due to reduced structural dynamics (Pagan et al., 2009). The glycans may also shield the hydrophobic regions of the protein and thereby increase the stability (Erbel et al., 1999; Pagan et al., 2009).

Metal-binding

Metal binding contributes to the stability and activity of many enzymes. Most α-amylases contain a conserved calsium ion, which is essential for stability and activity (reviewed by (Prakash and Jaiswal, 2010). The metal cofactor can stabilize the protein structure for example by coordinating, and thus stabilizing amino acids at the denaturation initiation site, as reported on *Bacillus licheniformis* xylose isomerase (Vieille *et al.*, 2001).

1.6 Modification of enzyme properties through protein engineering

The construction of modified proteins by site-directed mutagenesis was mainly used in the early years of protein engineering in the 1980s to study catalytic mechanisms and structure-function relationships by mutating the active sites of enzymes. Since then, the use of protein engineering has been extended to study for example stability, folding and specificity, through model systems reviewed by (Brannigan and Wilkinson, 2002). Understanding of the basis of the protein structure-function relatioship has opened possibilities to design enzymes for industrial applications, in which the process conditions are often very different from the biological environment of the enzyme. There are numerous successful studies in which for example stability, activity, pH dependence or functional expression of an enzyme has been altered using protein engineering techniques, exemplified by e.g. (Becker *et al.*, 2001; Bulter *et al.*, 2003; Morawski *et al.*, 2000; Song and Rhee, 2000; Wang *et al.*, 2005; Wohlfahrt *et al.*, 2003). Protein engineering strategies can be divided under two subtitles: rational design and directed evolution. In addition to these, there is a “semi-rational” strategy to engineer the protein stability, called the consensus concept. Regardless of the engineering technique used, there must be a possibility to produce the target enzyme (either in a homologous or heterologous expression host) to allow expression of the mutated gene. For rational enzyme design, detailed structural information or a structural model of the enzyme, combined with the biochemical data of the target enzyme, is also needed.

1.6.1 Rational mutagenesis

Site-directed mutagenesis is above all a tool to analyse the contribution of individual amino acids to a certain phenomenon. It is used to study the structure-function relationships of proteins, for example to reveal the catalytic mechanism. It can be used to create inactive mutants of enzymes in order to study the ligand-protein interactions in 3D complex structures. This involves mutagenesis of the potential active site residues and kinetic characterization of the mutants. The protein modification can be achieved also through exchange of whole domains or generation of fusion proteins, as the CBM fusions in this study. Detailed knowledge of the structure and function of the enzyme is used to design mutations to make the desired changes in properties. Analysis of the substrate binding and specificity can lead to engineering of enzymes with modified functions.

1.6.1.1 Rational mutagenesis for improved thermostability

The stability of the folded conformation of a protein depends on its primary structure, but in a complex manner, being determined by a multitude of both local and long-term interactions (see section 1.5.2 *Structural reasons for higher thermostability*). However, rational approaches to improve the stability of a folded protein conformation by site-directed mutagenesis have been used successfully. Because of the complexity of a protein fold, the chance of success is not easily predictable (Lehmann et al., 2000a). There are several examples of how enzyme thermostability has been improved by rational mutagenesis by introducing only one or two mutations (Mansfeld *et al.*, 1997; Matthews *et al.*, 1987; Watanabe *et al.*, 1994; Williams *et al.*, 1999), although in many cases single substitutions have only a relatively small effect. Since the effects of individual mutations are often additive (Akasako *et al.*, 1997; Shih and Kirsch, 1995), pronounced thermostabilization can be achieved by combining several individual mutants.

1.6.2 Directed evolution

Incomplete understanding of the structure-function relationships of enzymes and the relatively small number of available three-dimensional protein structures both limit the success of rational protein engineering by site-directed mutagenesis, but can be compensated by using directed evolution as a tool. In directed evolution approaches, random point mutations are introduced into a gene sequence, the

variant library is screened and improved variants are selected. These improved variants from the first round of mutagenesis can be used as templates for a second round of mutagenesis, or if several improved candidates are discovered they can be combined by gene shuffling.

One approach is to create single, random point substitutions by error-prone PCR with a non-proofreading DNA polymerase in the presence of $MnCl_2$ (Leung *et al.*, 1989) or by utilising nucleotide analogues that cannot be correctly read by the DNA polymerase (Zaccolo *et al.*, 1996). A critical point in random mutagenesis is careful tuning of the mutation frequency, since beneficial mutations are rare, whereas deleterious mutations are common (Harayama, 1998; Tracewell and Arnold, 2009). Although the combination of too many mutations in one clone may result in inactive enzymes, a too low mutation rate results in libraries predominated by the wild-type sequence. In general, 1.5 to 5 mutations per gene are often introduced (Miyazaki *et al.*, 2000; Morawski *et al.*, 2000).

In the second approach the randomization is targeted to certain areas within the gene. This targeting can be carried out based on structural information and/or previous knowledge about the function of the target protein. By targeting the mutagenesis the necessary library size can be reduced.

The third method is gene recombination of homologous genes or variants from the two first (above mentioned) approaches. This method of DNA shuffling (Stemmer, 1994) allows *in vitro* recombination of the different mutations from random mutagenesis rounds and combination of these beneficial mutations into one gene, greatly enhancing the process of directed evolution. In addition to combining mutations, DNA shuffling can eliminate neutral mutations introduced into the gene during random mutagenesis. The removal of so-called neutral mutations is important, as they may not be neutral under some other conditions which were not part of the selection/screening strategy. DNA shuffling has been extended from the original application to shuffle the gene variants from random mutant libraries to its use in family shuffling, where the sequence diversity is achieved by combining homologous and closely related genes by *in vitro* recombination (Crameri *et al.*, 1998). The original DNA shuffling method is based on fragmentation of DNA by DNAseI and reassembly of the fragments by PCR, where the fragments serve as primers in the PCR reaction. A drawback is that the shuffling reaction is biased due to unequal amplification and sequence dependent variations in DNAaseI digestion, and furthermore it depends on high homology (>70%) between the shuffled gene pools (Joern *et al.*, 2002).

Several other PCR-based methods have been described to overcome the above-mentioned weaknesses, including staggered extension process (StEP) (Zhao *et al.*, 1998), incremental truncation for the creation of hybrid enzymes (ITCHY) (Ostermeier *et al.*, 1999), family shuffling using single-stranded DNA (Kikuchi *et al.*, 2000), random chimeragenesis on transient templates (RACHITT) (Coco *et al.*, 2001) and Multiplex-PCR-based recombination (MUPREC) (Eggert *et al.*, 2005). In addition to the *in vitro* recombination methods, it has been shown that homologous genes can efficiently be recombined *in vivo* in the yeast *Saccharomyces cerevisiae* (Cherry *et al.*, 1999; Pompon and Nicolas, 1989). The recombination in yeast is based on the efficient repair of double-stranded breaks by intra-molecular homology-dependent recombination (Orr-Weaver *et al.*, 1981). The clear advantages of this method are that it is not limited by the insert size as is PCR, and that the recombined gene variants can be readily expressed in the shuffling host, allowing rapid screening. The weakness is that recombination events are not frequent; from one to three recombinations were detected by (Swers *et al.*, 2004).

Computational methods have been developed to predict intra-molecular sites suitable for recombination without interfering with the structural integrity of the target protein. The SCHEMA algorithm predicts which fragments of homologous proteins can be recombined without breaking the interactions in the three-dimensional structure of a protein (Voigt *et al.*, 2002). It has recently been used successfully to improve the properties of *Humicola* cellobiohydrolase of the GH-6 family. Almost 50% of the tested chimeras were active and 10% were more thermostable than any of the parents (Heinzelman *et al.*, 2009). However, the use of SCHEMA is limited to proteins for which 3D structural information is available.

1.6.2.1 Directed evolution for improved thermostability

As mentioned in previous the section (1.5), the stability of a protein results from a cooperative, balanced network of amino acids forming the protein fold. The fact that the general rules for thermostability are still unwritten supports the use of a random mutagenesis approach in the effort to create more stable biocatalysts. In addition, the results of directed evolution studies serve as a source of information about the molecular mechanisms of protein stability, and may help in targeting the rational mutagenesis strategies. There are numerous examples of studies in which directed evolution has been utilized. Esterase from *Bacillus subtilis* was stabilized by directed evolution, leading after six generations of mutagenesis to a

variant containing nine individual mutations and having a T_m improved by 14°C, while the catalytic activity at lower temperatures was retained (Giver *et al.*, 1998). The authors also showed that all the individual mutations contributed to the stability, and that the mutations were all located in the C-terminal domain of the protein, either on the surface or close to the surface of this domain. In another study, the thermostability of cold-adapted subtilisin S41 was clearly improved by iterative rounds of random mutagenesis, screening and DNA shuffling, without loosing the activity at low temperature (Miyazaki *et al.*, 2000). Although the calcium binding of the mutant was affected, none of the mutations were located in the calcium binding sites, and the authors concluded that the effect was due to long range interactions or subtle structural changes and was therefore not apparent from the homology model. Some of the mutations were located in loop areas, which cannot be studied accurately from a homology model, but the authors reasoned that a proline insertion (K211P) contributed to the stability by rigidifying the loop. DNA shuffling of the highly homologous endoglucanases EngB and EngD of *Clostridium cellulovarans* was used to improve the thermostability while retaining most of the activity (Murashima *et al.*, 2002). The authors also studied the locations of the mutations in the homology model of EngB. Since the mutations were clustered in certain surface areas of EngB, the authors speculated that they might stabilise the structure through ionic interactions. All these three examples show that although directed evolution can result in mutations which would have been difficult to predict even with high quality 3D structure, the outcome can be used to study the principles of thermostabilization.

1.6.3 Screening

In addition to creating diversity, an equally important and challenging step in directed evolution is screening or selection for the desired property. The outcome of directed evolution depends crucially on the screening method used. Directed evolution of large enzymes, even at a low mutation frequency, creates huge demands in detecting each mutation at least once, since for an average sized 300-amino acid protein there are 3.2×10^7 possibilities to make two random amino acid substitutions (Tracewell and Arnold, 2009). If there is a possibility to use a selection method instead of screening, the detection of improved variants from the library can be greatly enhanced. Selections enable examination of large numbers of variants (10^6–10^8), requiring that the improved enzyme imparts a

growth advantage to the cell (Arnold *et al.*, 2001). Selection is normally used for intracellular enzymes, for which the required link between cell survival and enzyme activity is easier to understand. Screening is required when the target activity cannot be linked to cell survival.

Here the question arises how many clones should be screened to find an improved variant. In practice, oversampling would be required to ensure that a particular variant has been examined during screening (Moore *et al.*, 1997). With realistic library sizes of large enzymes of 400–500 amino acids, this means that it is impossible to screen sufficient numbers of clones to be sure to find the best variant. In order to overcome this problem, random mutagenesis can be targeted for example to one subunit of the enzyme at a time (Otten *et al.*, 2002), or to regions in the vicinity of the substrate binding or active site (Dalby, 2003). The method used for screening is extremely important, since screening is often laborious and expensive, and in a worst case scenario can give misleading results. For example if the expression level between the mutated clones varies and only the enzyme activity is measured, the improved variants from the screen might turn out not to be improved in terms of activity but only in terms of expression. The expected improvements caused by single amino acid replacements are often also so small that the accuracy and sensitivity of the screen becomes crucial and the experimental errors should be very low. Enzyme activities can be screened on microtiter plates from culture supernatants or cell lysates, using assays that detect the formation of a coloured or fluorescent reaction product. Solid-phase screens are based on cleavage of the substrate on an agar plate or a filter and detection of the released product. This can be achieved for example by detecting a clearing zone or precipitation caused by enzymatic activity, or by formation of a fluorescent or coloured product (Murashima *et al.*, 2002; Wang *et al.*, 2005; Zhang *et al.*, 1997). Plate or filter screens suffer from the lack of quantitative detection of enzyme activities, and they are often used as a first, crude screening step to select the active mutants for further screening for example for improved stability (Murashima *et al.*, 2002). Digital imaging techniques have been used to detect halo formation using chromofor-labelled substrate (Hughes *et al.*, 2006) or release of a fluorescent group from fluorescent-labelled substrate (Joo *et al.*, 1999). Activity screens from liquid cultivations on microtiter plates (Bulter *et al.*, 2003; Cherry *et al.*, 1999; Miyazaki *et al.*, 2000) are more laborious and expensive but versatile, and the screening conditions can be tailored for the desired property. Robotics or automated pipetting stations can be used to boost the throughput. Screening capacity can be further increased by

assaying variants in pools, if the screen is sensitive enough to detect the desired activity even when the other variants in the pool are inactive (Hughes *et al.*, 2006). The screens in microtiter plates are also usually performed with fluorescent or colorimetric methods using substrate analogues, and therefore the results often also have to be verified for the actual substrate of interest.

1.6.3.1 Screening for improved thermostability

Screens for improved thermostability are most often based on measuring residual activity after a heating step. The strategy relies on the assumption that the more stable enzyme variant also has higher temperature optimum (see (Eijsink *et al.*, 2004) and references therein). Alternatively, the enzyme activity could be measured directly at high temperature. However, in practice the activity measurement at high temperature is often not possible in a high throughput manner due to technical limitations in the robotic setups. Another widely used screening strategy for thermostability is to measure the activity at two temperatures and monitor the ratio between the two numbers (Miyazaki and Arnold, 1999). By measuring the activity at low temperature as well, it is possible to monitor thermostability and activity in low temperatures simultaneously. This type of screen also helps to avoid variation caused by different expression levels.

1.6.4 Consensus concept for improved thermostability

Because rational design of thermostability has been shown to have a relatively low success rate, and directed evolution approaches depend on costly screening procedures, another useful method utilizing sequence statistics to localise stabilizing residues has been described (Lehmann et al., 2000b; Steipe et al., 1994). The consensus concept is based on aligning closely related homologous protein sequences for calculation of a consensus sequence. This method assumes that at each position of the protein sequence the most predominant amino acid substitution will lead to stabilization. For example, the consensus sequence was calculated from an alignment of 13 homologous phytases and the corresponding gene was synthesised. The protein product was 15–26°C more thermostable than any of the parent phytases (Lehmann et al., 2000b). Later, 6 more phytase sequences were added into the alignment, and recalculation of the concensus sequence led to an improved variant (7.4°C) (Lehmann *et al.*, 2002). When the additional substitutions were studied individually, it was discovered that 33% of

them were stabilizing and that most of the stabilizing mutations were located on the surface of the protein. A similar result was obtained when the consensus method was used to design stabilising mutations into a β-lactamase gene and a combinatorial library of the mutations was built and screened; here 25% of the mutations were stabilizing (Amin *et al.*, 2004). Contrary to the phytase study, the stabilizing mutations were distributed both on the surface and inside the protein fold, and no correlation between surface exposure and stabilization was detected.

1.7 Aims of the study

In this thesis the focus was on characterizing or creating new thermostable GH-7 family cellobiohydrolases suitable for hydrolysis of lignocellulose at high temperatures. Two cellobiohydrolases, Cel7B from the thermophilic fungus *Melanocarpus albomyces*, and Cel7A from the thermophilic fungus *Talaromyces emersonii*, were expressed in *Saccharomyces cerevisiae* to enable protein engineering studies. A second approach was to isolate new fungal GH-7 family cellulases and to characterize the properties of the purified enzymes. Cel7A enzymes from the fungi *Acremonium thermophilum, Chaetomium thermophilum* and *Thermoascus aurantiacus* were studied.

In particular, the following points were of major interest:

1) Directed evolution to improve the thermostability and activity of *M. albomyces* Cel7B

2) Rational mutagenesis to improve the thermostability and activity of *M. albomyces* Cel7B and *T. emersonii* Cel7A

3) Structure-function studies of novel, thermostable Cel7 cellobiohydrolases.

2. Materials and methods

A summary of the materials and methods used in this study is presented in this section. Detailed descriptions can be found in the original articles (I–IV).

2.1 Strains and vectors

Eschericia coli strain XL1-blue (Stratagene, USA) was used as a bacterial cloning host. *S. cerevisiae* strain NY179 (*leu* 2-3,112, *ura* 3-52) (received from Peter J. Novick, Yale University School of Medicine, USA) was used as a cloning and an expression host for the random mutant Cel7B library (I). *S. cerevisiae* strain INVSc1 (MATα, *his3-1, leu2, trp1-289, ura3-52*; Invitrogen, USA) was used as cloning and expression host for the *M. albomyces* Cel7B and *T. emersonii* Cel7A wild-type and mutant enzymes (II, III). A *T. reesei* strain (Vehmaanperä *et al.*, 1997), that lack the genes encoding the major cellulases Cel7A, Cel6A, Cel7B, and Cel5A was used as production host for *A. thermophilum* Cel7A, *C. thermophilum* Cel7A, *T. aurantiacus* Cel7A and *T. aurantiacus* + CBM fusion proteins (IV). A *T. reesei* strain (Haakana *et al.*, 2004) lacking the genes encoding the cellulases Cel6A, Cel7B and Cel5A was used as an expression host for *M. albomyces* Cel7B and the *Ma* Cel7B fusion protein containing the C-terminal CBM and linker from *T. reesei* Cel7A (II). The *S. cerevisiae* expression plasmid pYX212 (R&D Systems, USA) containing constitutive triose phosphate isomerase (TPI) promoter, a *polyA* terminator (from *Kluyveromyces lactis* HSF) and a *URA3* marker gene for selection was used as yeast expression vector for the *M. albomyces* Cel7B and *T. emersonii* Cel7A enzymes (I, II and III).

2.2 Construction of yeast expression plasmids (Publications I, II and III)

For expression of the *Ma cel7B* gene in *S. cerevisiae*, the cDNA encoding *M. albomyces* Cel7B (including the signal sequence) was isolated from the pALK1523 plasmid as a *Nde*I-*Eco*RI fragment and ligated into the yeast expression plasmid pYX212 (I). In Publication III the *T. emersonii* cel7A gene was amplified by PCR (Sambrook and Russel, 2001) with vector specific primers to generate 40 bp overhangs at both the 5´and 3´end of the genes, thus allowing cloning of the gene into the yeast expression vector pYX212 by *in vivo* yeast homologous recombination (Orr-Weaver *et al.*, 1981).

2.3 Site-directed mutagenesis

Mutations S290T and G4C+M70C were introduced into the *M. albomyces cel7B* gene and mutations G266C+D320C, T243C+A375C, Q190C+I200C, N54C+P191C, G4C+A72C into the *T. emersonii cel7A* gene by PCR using site-specific mutagenesis by overlap extension as described by (Sambrook and Russel, 2001). A detailed description of the mutagenesis and the primers used concerning *M. albomyces cel7B* and *T. emersonii cel7A* can be found in paper II and III, respectively. The mutated DNA was in each case cloned into the yeast expression vector pYX212 in *S. cerevisiae* by co-transforming the mutated PCR-product into yeast with a linearised pYX212 plasmid. The PCR products were generated by vector specific primers having 40–150 bp overhangs on both the 5´and 3´end, allowing cloning of the gene into the yeast expression vector pYX212 by *in vivo* yeast homologous recombination. *S. cerevisiae* transformations were carried out with a modified LiAc method using the Gietz Yeast transformation Kit (Tetra-Link, USA) according to the manufacturer's instructions. The selection of cellulose-positive transformants is presented in I, II and III.

The CBM fusions *Ma* Cel7B+*Tr* CBM, *Ta* Cel7A+*Tr* CBM and *Ta* Cel7A+*Ct* CBM were designed and constructed at Roal Ltd by genetically linking these one-module Cel7 proteins to the linker and CBM region of *T. reesei* Cel7A (= *Tr* CBM) or to the linker and CBM region of *C. thermophilum* Cel7A (= *Ct* CBM) (II and IV). Construction of *T. reesei* expression plasmids for production of the fungal Cel7 proteins from *T. aurantiacus* (*Ta* Cel7A, *Ta* Cel7A + *Tr* CBM, and *Ta* Cel7A + *Ct* CBM), *C. thermophilum* (*Ct* Cel7A) and *A. thermophilum* (*At* Cel7A and *At* Cel7B) is presented in (IV). The transformations were performed

at Roal Ltd as in (Penttilä *et al.*, 1987) with modifications described in (Karhunen *et al.*, 1993).

2.4 Random mutagenesis (Publication I)

The random mutant libraries (A and B) of Cel7B were generated by error-prone PCR using Mutazyme™ DNA polymerase according to GeneMorph PCR mutagenesis kit (Stratagene) instructions**.** The mutagenized PCR product was cloned into pYX212 plasmid (cut with *Eco*RI and *Xho*I) either by ligating the *Eco*RI and *Xho*I -digested PCR product with T4-ligase (library A) or making use of the homologous recombination in yeast by co-transforming the linearised vector with the PCR product directly into *S. cerevisiae* (library B) (Orr-Weaver *et al.*, 1981). The PCR product contained homologous flanking regions with the linearised vector at the 5' end (300 bp) and at the 3´ end (150 bp).

2.5 Screening of thermostable *M. albomyces* Cel7B variants (Publication I)

Robotic screening for thermostability was performed with MULac as substrate in 96-well plates. Thermostability was assessed based on residual activity measurement after a heat inactivation step of 4 minutes in 76°C. Robotic screening consisted of two rounds (screens 1 and 2), of which screen 2 was performed for the positive clones from screen 1 in order to verify the result. The robotic screening set up consisted of a QPix colony picker (Genetix, UK) to pick individual colonies into cultivation liquid (SC-Ura medium, pH 6) in 96-well microtiter plates, a Multimek 96 channel pipettor (Beckman, USA) to transfer the culture liquid after incubation for 3 days at 30°C into 2 separate assay plates, and a Multidrop dispenser to add the substrate for the activity assay. The activity towards MULac was measured from each clone twice; before and after the heat inactivation step. The release of MU in each well was measured with a Victor^2 V MTP reader (fluorescence ex. 355 nm and em. 460 nm). The ratios of the activities before and after the heat treatment were compared to those of the yeast-produced wild-type Cel7B. After screen 2, the clones which performed better than Cel7B wild-type were again picked to verify the result by a manual activity screen (3rd screen). The 3rd screen was performed by cultivating the yeast clones for 3 days in test tubes at 30°C, and measuring the MULac activity

(dublicate measurement) after 30 min reaction time, before and after heat treatment (for 5 and 10 minutes) at 70°C.

2.6 Protein production by *S. cerevisiae* in shake flasks (Publications I, II and III)

For production of the *Ma* Cel7B, *Te* Cel7A and the mutated variants, 2–5 l of SC-Ura medium was inoculated with 10% volume of an overnight pre-culture. The cultivation medium was buffered to pH 6 when *Ma* Cel7B was produced; in the case of *Te* Cel7A wild-type or mutant enzyme production non-buffered medium was used. The cultures were grown on a shaker (210 rpm) at 30°C. After 3 days of cultivation the supernatant was harvested by removing the cells by centrifugation. The details of the experimental procedures can be found in Publications I, II and III.

2.7 Protein production by *T. reesei* (Publications II and IV)

Shake flask and laboratory scale bioreactor cultivations of the heterologous Cel7-producing strains were performed at Roal Ltd as described in (Paloheimo *et al.*, 2003).

2.8 Purification of heterologously produced Cel7 enzymes (Publications I–IV)

The cellobiohydrolases produced in *S. cerevisiae* were purified by ion-exhange chromatography as described in Publications I, II and III. The cellobiohydrolases *At* Cel7A, *Ct* Cel7A, *Ta* Cel7A and *Ta* Cel7A+CBM fusion proteins, produced in *T. reesei*, were purified using a *p*-aminobenzyl 1-thio-β-cellobioside-based affinity column, prepared as described in (Tomme *et al.*, 1988a), and gel filtration chromatography (IV). The *T. reesei*-produced *Ma* Cel7B and *Ma* Cel7B+CBM were purified using ion-exchange and hydrophobic interaction chromatography methods as described in (III). The production and purification of the intact and catalytic (core) modules of *Tr* Cel7A were performed as described earlier (Suurnäkki *et al.*, 2000). The chromatography materials used in all the purifications were from GE Healthcare unless otherwise indicated. Fractions after each purification step were screened for the presence of Cel7B or Cel7A by measuring the activities against MULac and analysing them on 10% SDS-PAGE

and in Publications I and II by Western blotting, where Cel7B was detected with a polyclonal antibody. Fractions containing a single band of Cel7B or Cel7A on SDS-PAGE were pooled, concentrated and the buffer was changed to 50 mM NaAc pH 5.0 (or in the case of *Ma* Cel7B to 50 mM sodium phosphate buffer, pH 6.0).

2.9 Characterization of the purified proteins

Protein concentrations of the purified enzyme preparations were measured by their absorption at 280 nm using theoretical molar extinction coefficients, which were calculated from the amino acid sequences of each enzyme (Pace *et al.*, 1995). The molecular masses of the wild-type *Ma* Cel7B proteins expressed in *S. cerevisiae* and *T. reesei* were detected by electrospray ionization mass spectrometry (ESI-MS) at the University of Eastern Finland as described in (Jänis *et al.*, 2001) (II). The recombinant *Te* Cel7A protein was identified by mass mapping of the peptides from trypsin-digested protein and molecular masses of the Cel7A and deglycosylated Cel7A proteins were detected by MALDI-TOF mass spectrometry (Bruker Autoflex II) (III). The presence of the additional disulphide bridges in the *Ma* Cel7B and *Te* Cel7A disulphide-bridge mutants was confirmed by comparing the amount of free thiol in the enzyme samples to that in the wild-type enzyme (II and III). Free thiol was measured with Thiol and Sulfide Quantitation Kit (Molecular Probes, Invitrogen detection technologies, USA) according to the manufacturer's instructions. The correct processing of the N-terminus of the yeast-produced *Ma* Cel7B wild-type enzyme was verified by Edman degradation (I). Differences in the N-glycosylation pattern of the heterologously expressed *Ma* Cel7B were studied by EndoH treatment as described in Publication I, and those of *S. cerevisiae*-expressed *Te* Cel7A by EndoF treatment as described in III.

2.10 Enzyme activity assays

2.10.1 Soluble substrates

The specific activities of the purified Cel7 proteins towards the soluble 4-methylumbelliferyl-β-D-lactoside substrate MULac (Sigma) were measured as a function of temperature as described in (I, III and IV). pH optimum curves for the Cel7 enzymes were determined using MULac in 0.1 M McIlvaine -buffer (Dawson *et al.*, 1959), pH from 2 to 8, at 22°C as described in III and IV.

Kinetic constants (K_m and k_{cat} values) on MULac for *Tr* Cel7A, *Ta* Cel7A, *At* Cel7A and *Ct* Cel7A were assayed as described in IV, for *Ma* Cel7B and mutated variants as described in II and for *Te* Cel7A and mutated variants as described in III. Kinetic constants (K_m and k_{cat} values) and cellobiose inhibition constants (K_i) with 2-chloro-4-nitrophenyl-β-D-lactoside (CNPLac) were measured for *Tr* Cel7A, *Ta* Cel7A, *At* Cel7A and *Ct* Cel7A as described in IV and for *Ma* Cel7B and the mutated variants as described in II. The K_m and k_{cat} constants were calculated by fitting the data to the Michaelis-Menten equation using the program Microcal™ Origin®. Lineweaver-Burk plots, replots of their slopes against cellobiose concentration and Hanes plots were used to distinguish between simple competitive and mixed type inhibition and to determine the cellobiose inhibition constants (K_i). The corresponding α-value for mixed type inhibition was determined from a plot of the LWB intercept versus the cellobiose concentration.

2.10.2 Insoluble substrates

The microcrystalline cellulose (Avicel) hydrolysis assays were performed in different temperatures as described in I–IV. Phosphoric acid-swollen cellulose (PASC), prepared from Avicel according to (Walseth, 1952), was used as amorphous cellulose substrate as described in IV. The formation of soluble reducing sugars during Avicel and PASC hydrolysis was determined by the para-hydroxybenzoic acid hydrazide (PAHBAH) method (Lever, 1972) using a cellobiose standard curve (50 to 800 µM cellobiose). The amount of soluble cello-oligosaccharides (glucose – cellohexaose) released from the hydrolysis of polymeric substrates (Avicel and PASC) was determined (III and IV) by high-performance anion-exchange chromatography (HPAEC) as described in (Tenkanen *et al.*, 1997). The standards used were commercially available linear cello-oligosaccharides: glucose from Fluka (Switzerland), cellobiose, cellotetraose and cellopentaose from Serva (Germany), cellotriose and cellohexaose from Seikagaku (Seikagaku America, USA).

2.11 Thermostability measurements

Temperature-induced unfolding was monitored by tryptophan fluorescence by heating samples gradually (approximately 1°C /min) from 25°C up to 78°C and measuring the fluorescence intensity with a Varian Gary Eclipse spectro-

fluorometer (I). Intrinsic fluorescence of the samples was recorded after every 2°C by measuring emission at 340 nm using an excitation wavelength of 280 nm.

Circular dichroism (CD) measurements of the Cel7 and mutant enzymes were performed with a Jasco (model J-720) CD spectrometer equipped with a PTC-348WI Peltier-type temperature control system (Alder *et al.*, 1973). Spectra were recorded from 240 to 190 nm using a 1 mm cell and a bandwidth of 1 nm. The unfolding curves were measured at 202 nm using the temperature scan mode with a gradient of 2°C/min until a temperature of 80, 85 or 90°C was reached. Details of the measurements are presented in Publications I–IV. The effect of dithiothreitol (DTT) on the refolding of *Ma* Cel7B was studied in II.

The stabilities of *Ma* Cel7B produced in *T. reesei* and *S. cerevisiae* and of the three mutant enzymes, and of *Te* Cel7A and of the four mutants were determined as their half-lives during treatment at 70°C, by measuring their MULac activity after the heat treatment. The details can be found in Publications II and III.

2.12 Homology Modelling by the SWISS-MODEL and design of the mutations

The protein sequences of the catalytic modules without signal sequences were used for building three-dimensional models of the *At* Cel7A, *Ct* Cel7A and *Ta* Cel7A cellobiohydrolases by the automated SWISS-MODEL server (Schwede *et al.*, 2003). The templates and quality of the models are described in Publication IV. Swiss-PdbViewer (http://au.expasy.org/spdbv/) (Guex and Peitsch, 1997) and WHAT IF server (http://swift.cmbi.ru.nl/servers/html/) was used as tools to design the disulphide bridge mutations in the *Ma* Cel7B and the *Te* Cel7A wt structures.

3. Results

3.1 Heterologous expression of fungal cellulases in *S. cerevisiae* (Publications I–III)

To accommodate rapid and convenient expression for the mutagenesis studies, the single-module cellobiohydrolases *Ma* Cel7B from *M. albomyces* (I and II) and *Te* Cel7A from *T. emersonii* (III) were expressed in *S. cerevisiae* strains NY179 (I) and INVSc1 (II and III). In both cases, the fragment containing the cDNA of *Ma* Cel7B or *Te* Cel7A, including their own signal sequences, was cloned in pYX212 vector containing a *URA3* selection marker, under a strong constitutive TPI promoter. The proteins were produced in SC-ura medium, with a yield of about 3–5 mg of the protein per litre of culture supernatant. The heterologous expression of the enzymes was detected by measuring the activity on a soluble substrate, MULac, and by SDS-PAGE. The expression levels obtained were high enough to allow purification and characterization of the Cel7 enzymes without the need for further optimization of the growth conditions. Purification of the cellobiohydrolases from the yeast supernatants was performed with ion exchange chromatography (I–III), and activity measurements on soluble substrate MULac and insoluble substrate Avicel with both *Ma* Cel7B and *Te* Cel7A showed that the enzymes appeared to be expressed in yeast in a functional form. Only slight differences were detected when the wild-type enzymes from different host organisms were compared (II, III and (Tuohy *et al.*, 2002)). The yeast-expressed *Ma* Cel7B was visualised on SDS-PAGE (I, Fig. 1A, lane 5 and in II Fig. 2, lane 4). No major differences in the mobility of the heterologously expressed *Ma* Cel7B proteins were detected as compared with *Ma* Cel7B from the native host, despite the presence of two putative N-glycosylation sites at positions N5 and N320. Only a small fraction of the yeast-produced *Ma* Cel7B appeared to have glycans attached to it (I, Fig. 1A, lane 3), and electrospray ionization mass spectrometry

(ESI-MS) results (II) indicated that the purified *Ma* Cel7B did not contain any glycans incorporated by *S. cerevisiae*. Similarly to *Ma* Cel7B, *Te* Cel7A contains two putative N-glycosylation sites (N267 and N431). However, in this work yeast-produced *Te* Cel7A was clearly overglycosylated as compared to wild-type protein from the native host (Publication III, Fig. 2, Lanes 2 and 8). After deglycosylation with EndoF1 enzyme, the mobility of the yeast-produced *Te* Cel7A on SDS-PAGE was similar to that of the wild-type form purified from the native host (III, Fig. 2, Lane 3).

3.2 Production, purification and characterization of recombinant Cel7 proteins from *T. reesei* (Publications II and IV)

Cellobiohydrolase genes from *Acremonium thermophilum*, *Chaetomium thermophilum* and *Thermoascus aurantiacus*, in each case including their own signal sequences, were expressed in Roal Ltd under the strong *T. reesei cel7A* promoter in industrial production strains of *T. reesei*, in which genes encoding for the four major cellulases had been deleted, thereby facilitating purification and characterization of the recombinant Cel7 proteins (Vehmaanperä *et al.*, 2007). The Cel7 proteins were purified using an affinity column and gel filtration chromatography (IV). The gene coding for the *Ma* Cel7B, including its own signal sequence, was also cloned in Roal Ltd in a *T. reesei* expression vector under the *T. reesei cel7A* promoter and transformed into *T. reesei* strain A36 (Haakana *et al.*, 2004). *Ma* Cel7B from *T. reesei* was purified using ion exchange chromatography and hydrophobic interaction chromatography (II). The purified *Ma* Cel7B was analyzed on SDS-PAGE (II, Fig. 2, Lane 3), showing no difference in the mobility of the *Tr*-produced *Ma* Cel7B compared with Cel7B from the native host (Fig. 2, Lane 2) (Haakana *et al.*, 2004).

3.3 Sequence analysis

Comparison of the amino acid sequences of the five cellobiohydrolase enzymes showed high homology. They all belonged to glycoside hydrolase family 7, indicating that they all have similar protein fold and reaction mechanisms. The amino acid sequences coding for the mature proteins (without signal sequences) were aligned with the sequence of *T. reesei* Cel7A (Fig. 8) (Higgins *et al.*, 1996). The putative catalytic amino acids (E212, E217 and D214, in *Tr* Cel7A

numbering) are underlined in the sequence alignment. Since the Cel7 enzymes *Ma* Cel7B, *Te* Cel7A and *Ta* Cel7A consist merely of catalytic modules, the sequence identities were calculated without CBMs. The catalytic modules gave sequence identity values ranging from 50 to 81% (Table 2).

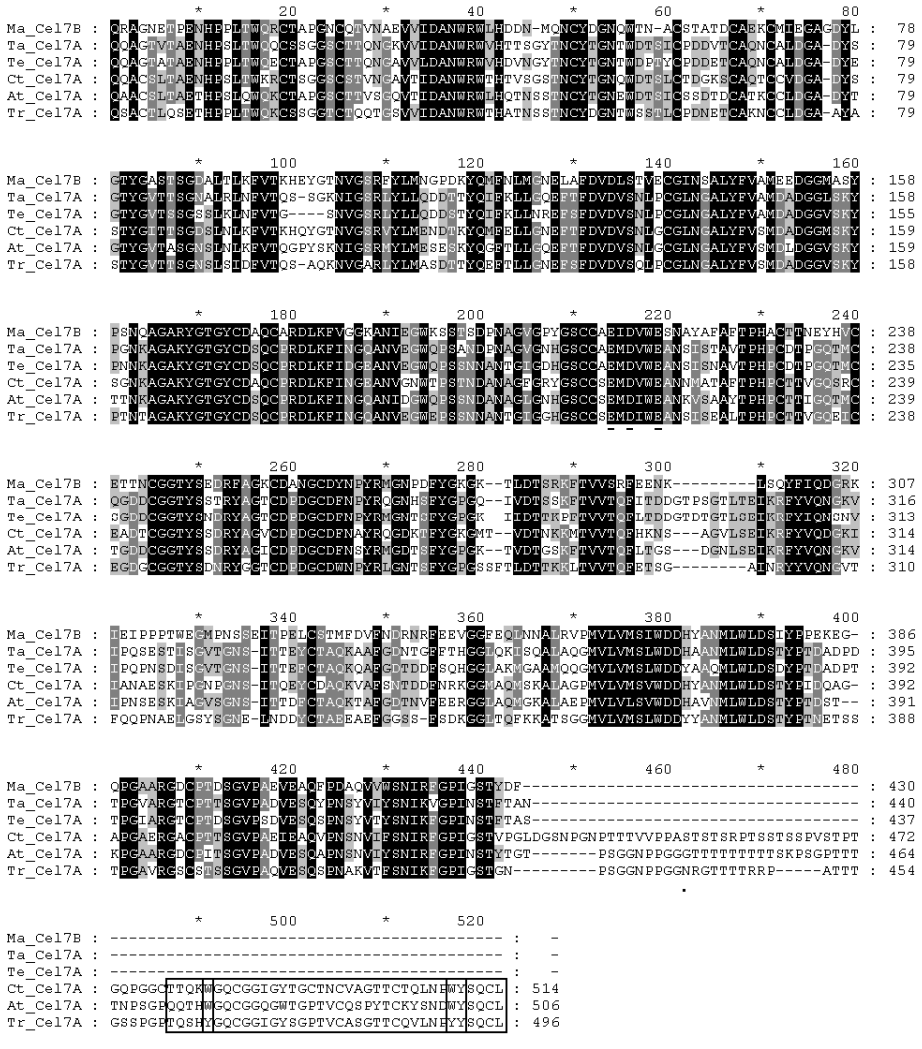

Fig. 8. Multiple sequence alignment of *M. albomyces* Cel7B, *T. aurantiacus* Cel7A, *T. emersonii* Cel7A, *A. thermophilum* Cel7A, *C. thermophilum* Cel7A and *T. reesei* Cel7A mature protein sequences. Residues in white against a black background are amino acids that are identical or have a conserved substitution in all six sequences. Residues in white against a grey background are amino acids that are identical or conserved in five out of the six sequences. The putative catalytic amino acids are underlined (E212, E217 and D214, in *Tr* Cel7A numbering) and the CBMs of *At* Cel7A, *Ct* Cel7A and *Tr* Cel7A are marked with a box.

Table 2. Sequence identities (%) of *M. albomyces* Cel7B, *T. aurantiacus* Cel7A, *T. emersonii* Cel7A, *T. reesei* Cel7A, *A. thermophilum* Cel7A and *C. thermophilum* catalytic modules.

	***At* Cel7A**	***Ct* Cel7A**	***Ma* Cel7B**	***Ta* Cel7A**	***Te* Cel7A**	***Tr* Cel7A**
***At* Cel7A**	100	71	56	73	70	64
***Ct* Cel7A**		100	58	68	64	61
***Ma* Cel7B**			100	54	53	50
***Ta* Cel7A**				100	81	66
***Te* Cel7A**					100	66
***Tr* Cel7A**						100

3.4 Mutagenesis

3.4.1 Random mutagenesis of *Ma* Cel7B and screening for higher thermostability (Publication I)

The complete gene (cDNA) sequence coding for *Ma* Cel7B was subjected to random mutagenesis by error-prone PCR using Mutazyme polymerase. Two libraries were created, having a total number of transformants of 2.6×10^5 clones in Library A, and 7.7×10^5 clones in Library B. Sequence analysis of 31 randomly picked clones showed that Library B contained on average 4.5 nucleotide mutations per gene, which corresponded with 2.2 amino acid changes per *Ma* Cel7B mutant.

These mutant libraries were screened for a cellulase with higher temperature stability using an activity assay based on the soluble fluorogenic substrate MULac, by measuring the residual activity of each clone after a heat inactivation step. The ratio of the activity in the heat-inactivated sample to that of the control sample was used as a measure for the thermostability of a mutant clone. Mutant clones having a higher activity ratio than wild-type Cel7B qualified for the next screening round. Altogether 14,580 colonies were screened with the robotic assay (screen 1), and 900 of these were analysed with the same robotic assay (screen 2). 200 positive clones were picked for a third, manual thermostability screening round at 70°C, and 49 clones were finally found to be more thermostable than the *Ma* Cel7B. The complete gene sequences of the 49 clones from screen 3 were determined, and showed a minimum of one and a maximum of four amino acid changes. After purification and characterization of the 9 best mutants (I), one mutant, containing a point mutation S290T, was found to be the most interesting for future mutagenesis studies. The mutant S290T had improved

thermostability as shown by T_m measurements (I, II). The activity towards crystalline substrate at high temperature (70°C) was clearly improved, presumably due to improved stability in the high temperature, as the activity on small substrate (MULac) was similar to that of the wild-type.

3.4.2 Site-directed mutagenesis to improve the thermostability and activity of the Cel7 enzymes

3.4.2.1 Design of disulphide-bridge mutants in *M. albomyces* Cel7B and *T. emersonii* Cel7A (Publications II and III)

Thermostability of *Ma* Cel7B and *Ta* Cel7A was improved by rigidifying their structures with additional disulphide bridges. The design of one of the disulphide bridges was based on a comparing the solved 3D structures with the *Tr* Cel7A's catalytic module structure (PDB code 7CEL). The latter has 10 disulphide bonds, whereas there are 9 S-S bonds in *Ma* Cel7B and *Te* Cel7A (Grassick *et al.*, 2004; Parkkinen *et al.*, 2008). This 10^{th} disulphide bridge was introduced at the equivalent position in *Ma* Cel7B and *Te* Cel7A structures by making two point mutations: G4C/M70C (*Ma* Cel7B; II, Fig. 1) or G4C/A72C (*Te* Cel7A; III, Fig. 1).

Several alternative disulphide mutants (III, Table 1) were designed in the *Te* Cel7A protein using Swiss-PdbViewer (http://www.expasy.ch/spdbv/) and WHAT IF disulphide-bridge predictor (http://swift.cmbi.ru.nl/servers/html/index.html) as tools to examine the Cel7A structure in order to plan the mutations and predict the possibility of disulphide bond formation. Mutants N54C/P191C, Q190C/I200C, T243C/A375C and G266C/D320C were designed to restrict the mobility of the surface loops of *Te* Cel7A (see Publication III, Fig. 6). After initial characterization of the yeast supernatants, it was found that three of the disulphide bridges described above increased the *Te* Cel7A activity towards MULac in high temperature (75°C) and they were combined to form a triple-mutant G4C/A72C/N54C/P191C/T243C/A375C.

3.4.2.2 Design of CBM fusions of *M. albomyces* Cel7B and *T. aurantiacus* Cel7A (Publications II and IV)

The hydrolytic activities on crystalline cellulose of the single-module cellobiohydrolases *Ma* Cel7B and *Ta Cel7A* were improved by genetically fusing

the linker and CBM of *Tr* Cel7A. In addition to these *Tr* CBM fusions, *Ta* Cel7A was also fused with *Ct* Cel7A CBM. The fusion proteins were produced in shake flask and laboratory scale bioreactor cultivations of the cellulose-negative industrial production strain *T. reesei* and purified for more detailed characterization in a similar manner to the wild-type *Ma* Cel7B and *Ta* Cel7A.

3.5 Activity measurements with soluble and insoluble substrates

3.5.1 Characterization of the activity on soluble substrates

All the Cel7 enzymes and mutated variants in this study exhibited Michaelis-Menten type kinetics on soluble MULac substrate, and the kinetic constants (K_m and k_{cat}) were determined at 22°C on MULac. Chromophoric substrate CNPLac was used to determine the K_i values for cellobiose inhibition in addition to the kinetic constants (K_m and k_{cat}). Table 3 compares enzyme properties of the Cel7 variants. The highest activity on MULac (k_{cat} value) of the cellulases characterized in this study, 69.0±3.2 min^{-1}, was observed with the *Ct* Cel7A and the lowest with *Ma* Cel7B (8.6 ± 0.2 min^{-1}), which was about 8-fold lower (Table 3).

The activity of Cel7 cellobiohydrolases is strongly inhibited by the hydrolysis product, cellobiose. Cellobiose inhibition for *Tr* Cel7A has been well studied, and strong competitive inhibition with an inhibition constant K_i around 20 μM and has been reported with CNPLac as a substrate (van Tilbeurgh and Claeyssens, 1985), whereas an almost 10-fold lower inhibition constant of 180 μM has been measured for *Te* Cel7A with $4NPG_2$ as substrate (Tuohy *et al.*, 2002). Here, we observed almost 20-fold differences between the cellobiose inhibition constants (K_i) of the wild-type cellobiohydrolases (Table 3), *Ma* Cel7A having the highest level of cellobiose inhibition (K_i = 6 μM) and *At* Cel7A the lowest (K_i = 141 μM).

Table 3. Comparison of the Michaelis-Menten, cellobiose inhibition constants of the Cel7A enzymes characterized in this study. The kinetic constants were measured on MULac (in pH 5.0) and on CNPLac (in pH 5.7) at 22°C. The assay conditions are presented in *Materials and methods.*

Enzyme	Variant (production host)	MULac			CNPLac					Ref
		k_{cat} (min^{-1})	K_M (µM)	k_{cat}/K_M ($min^{-1} M^{-1}$)	k_{cat} (min^{-1})	K_M (µM)	k_{cat}/K_M ($min^{-1} M^{-1}$)	K_i (Glc_2) (µM)	Type of inhibition	
Ma Cel7B	wt (*Tr*)	8.6 ± 0.2	230 ± 20	3.7×10^4	1.00 ± 0.05	620 ± 80	1.6×10^3	6.0 ± 2.0	Mixed, α = 1.5	1
	wt (*Sc*)	6.4 ± 0.3	280 ± 70	2.3×10^4	0.92 ± 0.04	800 ± 100	1.1×10^3	6.3 ± 2.0	Mixed, α = 1.1	1
	S290T (*Sc*)	6.5 ± 0.3	280 ± 40	2.3×10^4	1.23 ± 0.06	1160 ± 140	1.0×10^3	6.6 ± 2.0	Mixed, α = 1.4	1
	G4C/M70C (*Sc)*	5.0 ± 0.3	300 ± 50	1.6×10^4	1.09 ± 0.04	1000 ± 100	1.0×10^3	5.5 ± 2.0	Mixed, α = 1.2	1
	S290T/ G4C/M70C (*Sc)*	7.2 ± 0.4	300 ± 50	2.4×10^4	0.99 ± 0.06	750 ± 120	1.3×10^3	5.0 ± 2.0	Mixed, α = 1.9	1
Te Cel7A	wt (*Sc*)	31.1 ± 2.0	330 ± 70	9.4×10^4	nd	nd	nd	nd		2
	wt, deglycosy-lated (*Sc*)	33.2 ± 1.5	310 ± 40	1.1×10^5	nd	nd	nd	nd		2

Enzyme	Variant (production host)	MULac			CNPLac					Ref
	G4C/A72C (*Sc*)	29.7 ± 1.6	260 ± 50	1.1 x 10^5	nd	nd	nd	nd		2
	N54C/P191C (*Sc*)	35.9 ± 2.0	210 ± 50	1.7 x 10^5	nd	nd	nd	nd		2
	T243C/A375C (*Sc*)	26.8 ± 1.8	310 ± 70	8.6 x 10^4	nd	nd	nd	nd		2
	triple mutant (*Sc*)	26.7 ± 1.2	230 ± 40	1.2 x 10^5	nd	nd	nd	nd		2
At Cel7A	wt (*Tr)*	11.3 ± 0.7	220 ± 44	7.3 x 10^4	2.8 ± 0.1	2100 ± 150	1.3 x 10^3	141 ± 25	competitive	3
Ct Cel7A	wt (*Tr)*	69.0 ± 3.2	221 ± 35	3.1 x 10^5	19 ± 1	2000 ± 200	9.5 x 10^3	39 ± 14	competitive	3
Ta Cel7A	wt (*Tr)*	19.5 ± 0.8	268 ± 38	7.3 x 10^4	1.7 ± 0.1	990 ± 70	1.7 x 10^3	107 ± 14	competitive	3
Tr Cel7A	wt (*Tr)*	28.7 ± 1.0	287 ± 31	1.0 x 10^5	2.6 ± 0.05	520 ± 30	5.0 x 10^3	19 ± 4	competitive	3

nd = not determined

References: 1. Publication II, 2. Publication III, 3. Publication IV

In addition to the kinetic constants, MULac was used to measure the activity as a function of pH and temperature. The pH optimum for *At* Cel7A, *Ct* Cel7A, *Te* Cel7A and *Ta* Cel7A determined in McIlvaine buffer (from pH 2 to 8) using MULac as the substrate at 22°C was in the range pH 4 to 5 (III, IV), while *Ma* Cel7B demonstrated higher pH optimum of 6 (unpublished result). The temperature activity profiles (10 min assays with the soluble substrate, MULac) of the purified Cel7 enzymes (*Ma* Cel7B, *Te* Cel7A, *At* Cel7A, *Ct* Cel7A, *Ta* Cel7A and *Tr* Cel7A) and of the mutated variants were determined at pH 5.0 (except *Ma* Cel7B at pH 6.0) between 30 and 80°C.

No clear differences were detected between the wild-type *Ma* Cel7B and the mutants from the random mutagenesis study in terms of MULac temperature optimum (I, Fig. 3). The results of the rational mutagenesis study (*Ma* Cel7B, II) on the other hand, showed that the MULac temperature optima of the disulphide mutants had been shifted ~5°C higher, indicating a stabilizing effect of the additional S-S bridge. The activities of the *Te* Cel7A mutants as a function of temperature were also altered when compared to the wild-type, as shown in III, Fig. 5A and B. The mutant N54C/P191C exhibited clearly higher activity but approximately the same temperature optimum as the wild-type, 65°C, whereas the other mutants showed improved activities only in higher temperatures (75–80°C). The temperature optima of cellobiohydrolases *At* Cel7A, *Ct* Cel7A and *Ta* Cel7A were close to each other, varying from 60 to 65°C (IV, Fig 3). All in all, the most active Cel7 enzyme in this study was *Ct* Cel7A (opt. 65°C, rate 226 ± 1 min^{-1}) (IV) and the second most active was the *Te* Cel7A mutant N54C/P191C (opt. 65°C, rate 202 ± 15 min^{-1}) (III).

3.5.2 Characterization of the activity on insoluble substrate

The activity of the purified Cel7 enzymes (*Ma* Cel7B, *Te* Cel7A, *At* Cel7A, *Ct* Cel7A, *Ta* Cel7A and *Tr* Cel7A) and the mutated variants towards microcrystalline cellulose Avicel was determined at pH 5.0 (except *Ma* Cel7B at pH 6.0) in different temperatures. The activities of the thermostable mutants of *Ma* Cel7B were measured at elevated temperatures 70 and 75°C. The *Ma* Cel7B S290T mutant can hydrolyse crystalline cellulose at 70°C 2-fold more effectively than the Cel7B wt enzyme (II, Fig. 7a). At 75°C both *Ma* Cel7B mutants with an additional disulphide bridge (G4C/M70C and G4C/M70C/S290T) were clearly more active than the Cel7B wt or the S290T mutant, whereas the activities were similar to *Ma* Cel7B wt at 70°C (II, Fig. 7b), indicating that the improved thermostability also improves the

activity at elevated temperatures. In the case of *Te* Cel7A, the disulphide bridge mutant N54C/P191C, which showed higher activity than the wild-type on soluble substrate, was also more active on Avicel at 70°C than wild-type or the other mutants. However, at 75°C, both wild-type and N54C/P191C lost most of their activity and the other disulphide bridge mutants hydrolysed Avicel more efficiently. At 80°C, the only active enzyme in this study was the *Te* Cel7A triple-mutant (Publication III, Figs. 5A-C).

The presence of a family-1 CBM increased the activity of all enzymes studied here as native 2-module versions (*Ct* Cel7A and *Tr* Cel7A) or fusion constructs (*Ma* Cel7B+*Tr* Cel7A CBM, *Ta* Cel7A+*Tr* Cel7A CBM and *Ta* Cel7A+*Ct* Cel7A CBM) (Publications II and IV). The Avicel activity results of each enzyme are shown in more detail in Publications I–IV.

The hydrolysis of amorphous cellulose PASC in 45°C with the 2-module versions of *At* Cel7A, *Ta* Cel7A +*Ct*/*Tr* CBM and *Tr* Cel7A is shown in Publication IV, Fig. 5. In accordance with the Avicel hydrolysis studies, *Ct* Cel7A also had the highest activity on PASC. In general, PASC is a better substrate for all the tested cellulases. Contrary to Avicel solubilization, hydrolysis of amorphous cellulose is not invariably accelerated by the presence of CBM (unpublished results). The activities with the 2-module versions of *Ct* Cel7A and *Ma* Cel7B on PASC were slightly higher than the activities with the corresponding catalytic modules, whereas the two-module versions of *Ta* Cel7A and *Tr* Cel7A had similar or lower activity on PASC than the corresponding catalytic module.

3.5.3 Comparison of activities on soluble and insoluble substrates

To compare the activities of the one- and two-module versions of *At* Cel7A, *Ct* Cel7A, *Ma* Cel7B, *Ta* Cel7A and *Tr* Cel7A on soluble and insoluble substrates in different temperatures, the activities on Avicel at 45°C and 70°C and on MULac at 40°C and 70°C were plotted in the same graph (Fig. 9). The comparison revealed differences in the temperature behaviour of the thermostable enzymes. *Ct* Cel7A had high MULac activity but its Avicel activity was clearly diminished at high temperature. On the other hand, *Ta* Cel7A + $CBM_{CtCel7A}$ had retained its activity towards Avicel at elevated temperature. The MULac activities were measured during a short time period of 10 min, whereas Avicel hydrolysis results were from a long incubation of 17 h. Therefore differences in the temperature profiles could be due different stabilities during longer high temperature exposure. However, the unfolding temperatures (T_m) of these

enzymes, measured by CD (see section Thermostability), do not indicate any differences in the temperature stability of *Ct* and *Ta* Cel7A. This divergent behaviour with soluble and insoluble substrate might indicate differences in how these enzymes are stabilised by the substrate, or differences in product inhibition as a function of temperature.

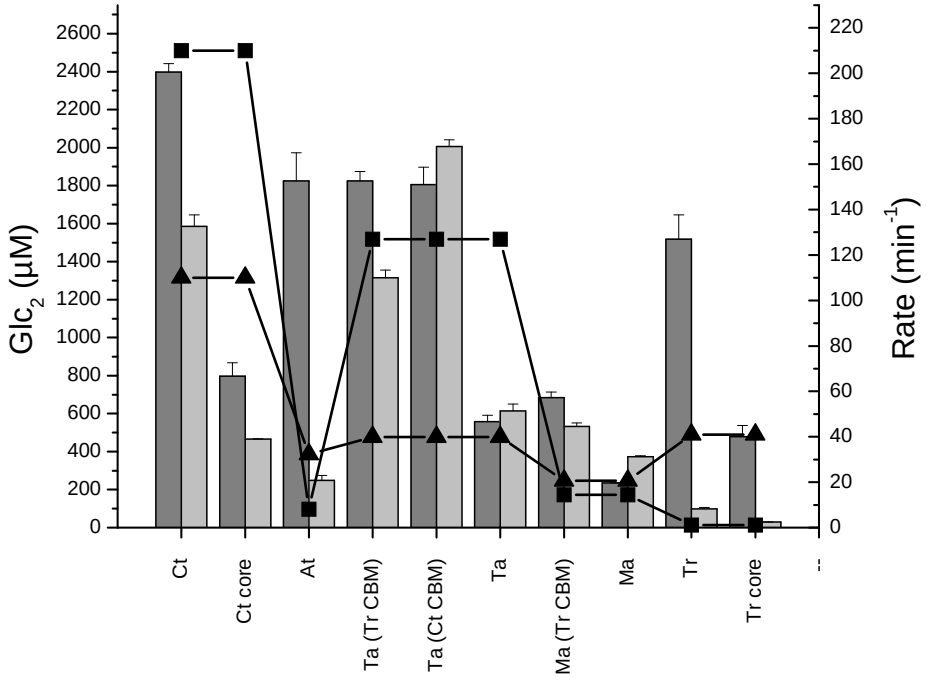

Fig. 9. Degradation of Avicel and MULac by the one- and two-module versions of *T. reesei*-produced *Ct* Cel7A, *At* Cel7A, *Ta* Cel7A, *Ma* Cel7B, and *Tr* Cel7A. 1% Avicel (w/v) was mixed with 1.4 µM enzyme in 50 mM NaAc pH 5 or in 50 mM sodium phosphate pH 6 for *Ma* Cel7B. The soluble reducing sugars were measured after 17 h Avicel hydrolysis in 45°C (gray bar) and 70°C (light gray bar) with PAHBAH reagent using cellobiose as a standard and calculated as micromolar cellobiose released (Left Y-axis). MULac activity was measured at 40°C (▲) and 70°C (■) (Right Y-axis), using 0.12 µM enzyme and 3.2 mM substrate in pH 5. In both temperatures, the initial rates were determined during the first 10 min by taking samples at designated time points and stopping the reaction with 0.5 M Na_2CO_3, after which the fluorescence of the samples was measured and calculated from the standard curve as described in Materials and Methods. All data are averages of duplicate assays, with the range between duplicates shown as error bars.

3.6 Thermostability

The unfolding temperatures of all the purified proteins in this study were measured by CD spectroscopy (I–IV). The results are summarised in Table 4. The results show clearly that the *Te* Cel7A mutants were the most thermostable enzymes in this study, the triple-mutant having a T_m of 84°C. The residual activities (at 22°C) of the purified Cel7 proteins were measured after different incubation times at 70°C to determine the half-lives ($T_{1/2}$) of the enzymes (Table 4). The experiment was performed in the presence of 1mM DTT. Similarly to the T_m measurement experiment showed that the *Te* Cel7A triple-mutant was more stable than the other enzymes studied here. Cellobiohydrolases from GH-7 family (Cel7) have been shown to refold after thermal treatment and regain most of their activity back (Publications I to IV and (Boer *et al.*, 2000)). This has been attributed to the existence of the nine (or ten) S-S bridges in the catalytic modules of all known GH-7 family wild-type cellobiohydrolases. Due to the refolding, the residual activities of Cel7 cellobiohydrolases after heat treatment cannot be measured reliably without DTT (II). It was shown by CD experiments that 1mM DTT (supposed to break the S-S bridges at high temperature (Scigelova *et al.*, 2001)) prevents the refolding after heat treatment, but does not affect the CD spectra or the activity at room temperature. Therefore, the addition of 1 mM DTT was used in the residual activity measurements (II and III). Table 4 shows that the measured unfolding temperatures (without DTT) and the half-life times (with DTT) for the different protein variants appeared to correlate. However, the possibility that the half-life time measurements would reflect (at least partially) the stabilising effect of the additional cysteines as such and not the disulphide bridges cannot be ruled out.

Table 4. Comparison of the unfolding temperature (T_m) and half-life time $T_{1/2}$ of the Cel7 enzymes characterized in this study. T_m was measured by CD spectroscopy. $T_{1/2}$ was estimated by measuring the residual activity after incubation in 70°C in the presence of 1 mM DTT. The assay conditions are presented in *Materials and methods.*

Enzyme	Variant (production host)	Unfolding temperature (T_m) (°C)	Half-life time $T_{1/2}$ in 70°C (min)	Ref
M. albomyces Cel7B	wt (*Tr*)	68.0 ± 1	Nd	1
	wt+CBM (*Tr*)	70.5 ± 1	Nd	1
	wt (*Sc*)	68.5 ± 1	4	1
	S290T (*Sc*)	70.0 ± 1	4	1
	G4C/M70C (*Sc)*	71.0 ± 1	13	1
	S290T/G4C/M70C (*Sc)*	72.5 ± 1	19	1
T. emersonii Cel7A	wt (*Te*)	74.0 ± 1	Nd	2
	wt (*Sc*)	75.0 ± 1	30	2
	wt, deglycosylated (*Sc*)	74.5 ± 1	15	2
	G4C/A72C (*Sc*)	80.0 ± 1	270	2
	N54C/P191C (*Sc*)	78.5 ± 1	150	2
	T243C/A375C (*Sc*)	79.0 ± 1	90	2
	triple mutant (*Sc*)	84.0 ± 1	320	2
A. thermophilum Cel7A	wt (*Tr)*	69.0 ± 1	Nd	3
C. thermophilum Cel7A	wt (*Tr)*	75.0 ± 1	21	3, 4
	wt, catalytic module papain digest (*Tr)*	71.0 ± 1	Nd	4
T. aurantiacus Cel7A	wt (*Tr)*	75.0 ± 1	31	3, 4
	wt + *Ct* CBM (*Tr)*	75.0 ± 1	Nd	3
	wt + *Tr* CBM (*Tr)*	75.0 ± 1	Nd	3
T. reesei Cel7A	wt (*Tr)*	65.0 ± 1	Nd	3
	wt, catalytic module (*Tr)*	67.0 ± 1	Nd	4

nd = not determined

References: 1. Publication II, 2. Publication III, 3. Publication IV, 4. Sanni Voutilainen, unpublished

4. Discussion

The aim of this work was to study the structure-function relationship of thermostable GH-7 family cellobiohydrolases and improve their stabilities and activities through protein engineering. Altogether five fungal Cel7 cellobiohydrolases derived from thermophilic, ascomycete fungi, were studied. *T. reesei* Cel7A was used as a reference enzyme. Three of these cellulases exist in nature as single module enzymes (*Ma* Cel7B, *Ta* Cel7A and *Te* Cel7A) and three as 2-module enzymes having a catalytic module and a cellulose-binding module (*Ct* Cel7A, *At* Cel7A, *Tr* Cel7A).

4.1 Heterologous expression of cellobiohydrolases in *S. cerevisiae* and effects of N-glycosylation

The thermostability and activity of the single module enzymes *Ma* Cel7B and *Te* Cel7A were improved by using two different types of protein engineering methods; random mutagenesis and site-directed mutagenesis. Both mutagenesis strategies require a suitable protein expression method and especially random mutagenesis relies on expression hosts with a high transformation frequency. The *Ma cel7B* and *Te cel7A* genes were expressed in *S. cerevisiae*, since it has high enough transformation frequency (up to 10^6) and no endogenous cellobiohydrolase activities. Both Cel7 cellulases were expressed in the yeast with their own signal sequences, under a constitutive TPI promoter (Publications I, II and III). Although *E. coli* would have been an ideal host for directed evolution and other protein engineering purposes due to fast and convenient generation of mutants, *S. cerevisiae* was chosen as host, since the only published trials with *E. coli* have been reported to be unsuccessful: *T. reesei* Cel7A as well as *T. emersonii* Cel7A have been produced as inclusion bodies (Grassick *et al.*, 2004; Laymon *et al.*, 1996).

One of the disadvantages of *S. cerevisiae,* as compared to the native filamentous fungal hosts, is its tendency to hyperglycosylate the extracellularly produced proteins by adding long mannose chains (up to 200 mannose residues) to the asparagine residues on the N-X-S/T recognitions sites (see also below). N-Glycosylation is a co-translational protein modification that can affect protein structure. Protein folding may be altered or facilitated by addition of the carbohydrate to a partially folded nascent polypeptide, and the carbohydrate can also alter the properties of the mature protein (Imperiali and O'Connor, 1999). Several different heterologous expression hosts, including *S. cerevisiae, Pichia pastoris, Aspergillus oryzae, A. niger* and insect cells have been used for expressing fungal cellobiohydrolases and variable glycosylation has been reported to be a major limitation. Heterologous expression of *T. reesei* cellobiohydrolase Cel7A in *S. cerevisiae* and *P. pastoris* has been reported to suffer from low production levels and extensive over-glycosylation leading to reduced hydrolytic activity on polymeric substrates (Boer *et al.*, 2000; Penttilä *et al.*, 1988; Reinikainen *et al.*, 1995). Interestingly, the expression of *T. reesei* cellobiohydrolase Cel6A in *S. cerevisiae* has been more successful with 100 mg/l expression level, and only slightly reduced activity as compared to Cel6A produced in the native host (Penttilä *et al.*, 1988). *Tr* Cel7A expression has also been demonstrated in insect cells (*Trichoplusia ni*), leading to a reasonable expression level of 20 mg/l, but again slightly reduced activity as compared to Cel7A produced in *T. reesei* apparently due to differences in glycosylation and/or processing of the recombinant enzyme (von Ossowski et al., 1997). Similar results (the recombinant enzyme having lower activity than the native Cel7A) have also been obtained with the filamentous fungi *Aspergillus oryzae* and *A. niger*, once more the explanation being glycosylation diverging from the native enzyme (Jeoh *et al.*, 2008; Takashima *et al.*, 1998). The N-glycosylation of *T. reesei* Cel7A from the native host can also vary considerably merely due to the cultivation conditions of the fungus. Three out of the four potential N-glycosylation sites (N45, N270 and N384) of the catalytic module are in fact glycosylated. Depending on the fermentation conditions these glycans are processed/hydrolysed unequally during the cultivation (Stals *et al.*, 2004). *T. reesei* Cel7A has been shown to be also O-glycosylated by the native host. O-Glycosylation has been detected in the linker between the catalytic module and the CBM (Harrison *et al.*, 1998). All the linker threonine and serine residues were found to be O-glycosylated with one to three mannose residues (Harrison *et al.*, 1998). On the contrary to N-glycosylation, O-glycosylation in *S. cerevisiae* differs relatively little in size from *T. reesei*, since O-glycans in

S. cerevisiae are linear chains of five mannose residues (Tanner and Lehle, 1987). Considering the complications that may be caused by differences in the N-glycosylation, understanding of this complex phenomenon would provide essential information for future protein production. Especially current efforts to express high levels of cellulases in *S. cerevisiae* for one-step conversion of crystalline substrate to ethanol would benefit from detailed knowledge of the causes and effects of hyperglycosylation. This consolidated bioprocessing (CBP) approach is considered to be one of the most promising strategies for commercial cellulosic bioethanol production (Banerjee et al., 2010; Lynd et al., 2002; Lynd et al., 2005).

Protein expression of both the *Ma cel7B* and *Te* Cel7A genes in *S. cerevisiae* was successful in terms of production level; the concentration of Cel7 proteins in the culture supernatant was moderate (3–5 mg/l), but sufficient to allow initial characterization directly from the supernatants (Publications I, II and III). This latter aspect is particularly important for random mutagenesis studies, which involve screening of thousands of variants and requires reliable measurement of the activity from the supernatants without concentration or purification steps. Purification of the wild-type *Ma* Cel7B and *Te* Cel7A enzymes and the selected variants for a more detailed characterization could also be conveniently carried out from shake flask cultivations of the yeast clones.

The properties of the yeast-produced cellulases were compared to those expressed in *T. reesei* (*Ma* Cel7B), or in the native hosts *Melanocarpus albomyces* (*Ma* Cel7B) and *Talaromyces emersonii* (*Te* Cel7A). The yeast-produced *Ma* Cel7B resembled the *T. reesei*-produced *Ma* Cel7B closely, showing almost identical properties (e.g. T_m and catalytic constants on soluble substrates) (Tables 3 and 4). Interestingly, only a minor fraction of the yeast-produced *Ma* Cel7B appeared to be glycosylated (Publications I and II), despite the two putative N-glycosylation sites (N5 and N320) present in and located on the surface of the *Ma* Cel7B catalytic module. Furthermore, no N-glycans could be detected in the *Ma* Cel7B 3D structure of (Parkkinen *et al.*, 2008), solved using *T. reesei*-produced *Ma* Cel7B protein. Similarly to *Ma* Cel7B, *Te* Cel7A could be expressed in *S. cerevisiae* in a functional form, showing similar specific activity on MULac and thermostability (T_m) to that of the native Cel7A expressed in *T. emersonii* (Publication III). However, contrary to *Ma* Cel7B, *Te* Cel7A was clearly overglycosylated by yeast, leading to apparent increase of molecular weight by 4.6 kDa as judged from MALDI-TOF MS results. All three mutants seemed also to be overglycosylated in a similar manner (see SDS-PAGE in III, Fig. 2.). *Te* Cel7A

has two N-glycosylation sites on the catalytic module (N267 and N431), but located on different places as compared to *Ma* Cel7B. N-acetylglucosamine (GlcNAc) residues have been detected on both of them in the 3D structure (Grassick *et al.*, 2004).

The consensus sequence for an N-glycosylation site is N-X-S/T (were X is any amino acid except proline), but not all sites are N-glycosylated, since in addition to the recognition sequence neighbouring amino acids as well as the local 3D structure also play a role in recognition by the oligosaccharyl transferase that catalyzes asparagine-linked N-glycosylation (Yan and Lennarz, 2005). The local 3D structure around the N-glycosylation sites in *Ma* Cel7B does not give any indication of why they would not be glycosylated. In fact both the N-glycosylation sites are located in loop areas, which in general are considered preferable for N-glycosylation (Imberty and Perez, 1995), although a more recent study has shown that N-glycosylation can occur on all types of secondary structures (Petrescu *et al.*, 2004). In the case of *Te* Cel7A one of the N-glycosylation sites (N267) is located just before an alpha helix, which possibly increases the probability of N-glycosylation, as junction points of secondary structure elements have been found to be more prone to be occupied (Petrescu *et al.*, 2004).

It has been shown in a statistical analysis that the probability of an N-glycosylation site being occupied is slightly higher when the +2 position of the recognition sequence is threonine instead of serine (Petrescu *et al.*, 2004). Both *Ma* and *Te* Cel7 enzymes have one N-glycosylation site with threonine and one with serine in position +2. The amino acid sequence around the N-glycosylation sites N5 and N320 in *Ma* Cel7B have elements that reduce the probability of N-glycosylation, e.g. acidic amino acids in close proximity upstream from the N-glycosylation sites (Petrescu *et al.*, 2004). Both *Te* Cel7A N-glycosylation recognition sites show elements that are expected to increase the probability of N-glycosylation, i.e. bulky, hydrophobic amino acids at position +3. The occurrence of non-polar residues in the -1 and -2 positions (here both in *Ma* Cel7B and *Te* Cel7A sequences) also increases the N-glycosylation probability. These slight differences may provide an explanation for the divergence seen in the N-glycosylation of the two yeast-expressed Cel7 cellobiohydrolases, but more thorough examination and site-specific mutations would be required to clarify the role of the individual amino acids near the N-glycosylation sites.

Both the *Ma* Cel7B and *Te* Cel7A are single-module cellulases, which do not contain a CBM. When expressed in *S. cerevisiae* these catalytic modules were shown not to contain O-glycans since their molecular masses (detected with MS

and gel electrophoresis) after purification or enzymatic deglycosylation were nearly identical to their calculated molecular masses (II and III). This is in accordance to earlier studies of *T. reesei* Cel7A, which has been shown to be O-glycosylated only in linker region connecting the catalytic module and CBM (Harrison *et al.*, 1998). In the case of the two-module enzymes produced in *T. reesei* (Publications II and IV) it is likely that the linker regions are O-glycosylated with one to three mannose residues through threonine and serine residues similarly to *T. reesei* Cel7A (Harrison *et al.*, 1998). The effect of these glycans to the properties of the different cellulases in this study remain unclear, because the addition of the linker and the CBM seemed to stabilize *Ma* Cel7B by 2.5°C, but T_m of *Ta* Cel7A remained unaltered, while the T_m value of the catalytic module of *Tr* Cel7A was higher than that of the full-length enzyme (Table 4).

4.2 Structure-function relationships and overall comparison of the six Cel7 catalytic modules

As shown in Table 2, the overall sequence identities of the six Cel7 catalytic modules studied vary between 50–81%, *Ta* Cel7A and *Te* Cel7A sharing the highest amino acid identity, and *Tr* Cel7A and *Ma* Cel7B the lowest. The 3D structures of the catalytic modules of three cellobiohydrolases in this study (*Ma* Cel7B, *Te* Cel7A and *Tr* Cel7A) have been solved and they all showed a similar, GH-7 family cellobiohydrolase fold (Divne *et al.*, 1994; Grassick *et al.*, 2004; Parkkinen *et al.*, 2008). The catalytic modules are composed in each case of two antiparallel β-sheets that stack onto each other to form a β-sandwich. The substrate binding and catalytic site is in a 50Å long tunnel formed by altogether 6 loops (loops 1–6, III, Fig. 1) protruding from the β-sandwich. The overall structure is stabilized by nine (*Ma* Cel7B and *Te* Cel7A) or ten (*Tr* Cel7A) disulphide bridges, which are located between conserved residues. The complex structures of *Ma* Cel7B and *Tr* Cel7A have shown that the substrate-binding tunnel contains 9–10 glycosyl binding sites, numbered from -7 to +2 or +3 (Divne *et al.*, 1998; Parkkinen *et al.*, 2008). The glycosyl binding sites are lined with amino residues that form hydrogen bonds with the sugar units as well as conserved tryptophan residues at subsites -7, -4, -2, and +1, which provide stacking interactions with the sugar rings. The catalytic site is located between subsites -1 and +1, where the conserved catalytic residues E212 and E217 (in *Tr* Cel7A numbering) are located approximately on opposite sites of the cleaved glycosidic linkage. In addition to the nuclophile (E212) and proton donor

(E217), D214 is assumed to assist the hydrolysis by controling the correct positioning and the protonation state of E212. The roles of the three conserved carboxylic acid residues have been studied by site-directed mutagenesis of *Tr* Cel7A (Divne *et al.*, 1998; Ståhlberg *et al.*, 1996). These three amino acids are in similar positions in *Ma* Cel7B and *Te* Cel7A and are proposed to function in a similar manner.

The homology models built for the *At* Cel7A, *Ct* Cel7A and *Ta* Cel7A catalytic modules (Publication IV) share the main structural features of the above-mentioned Cel7 cellobiohydrolase structures. The amino acid residues lining the active site tunnel are highly conserved and the three carboxylic acid residues around the putative catalytic site are in each case conserved.

4.2.1 Differences detected in the Cel7 catalytic module structures and function

Concerning the six loop structures forming the active site tunnel of Cel7 cellobiohydrolases, there was relatively little variation in the lengths of the loop structures forming the active site tunnels of the six Cel7 enzymes studied, but somewhat more variation in the amino acid compositions of the loops. The biggest differences were observed at the entrance of the substrate binding tunnel (see residues 99 to 104 (*Tr* Cel7A numbering) in the sequence alignment, Fig 8). This loop 2 is conserved in length in *At, Ct* and *Ma* Cel7 and has one residue insertion when compared to *Ta* and *Tr* Cel7A and a four residue insertion when compared to *Te* Cel7A. In particular this loop area of *Ct* Cel7A is almost identical in *Ma* Cel7B, containing only one conservative amino acid substitution. The difference in the loop length is expected to change the geometry of the tunnel entrance around the -7 subsite, with the result that *At, Ct* and *Ma* enzymes have a more closed entrance area than *Ta* Cel7A, *Tr* Cel7A and especially *Te* Cel7A. Furthermore, *Ta* and *Te* Cel7A have a tyrosine residue (Y47) located close to the entrance, opposite to loop 2, possibly creating an additional substrate binding subsite, -8, as suggested for *P. chrysosporium* Cel7D (Munoz *et al.*, 2001). Tyrosine residue Y100 at the tip of the loop 2 in *Ma* Cel7B, which is also seen at a similar position in *At* and *Ct* Cel7A structural models, has been proposed to have similar role of guiding of the cellulose chain into the tunnel as that of the residue Y47 in *P. chrysosporium* Cel7D (Parkkinen *et al.*, 2008).

There are also structural differences at the other end of the tunnel, near the product site, which might be related to differences detected in the end-product

(cellobiose) inhibition (Table 3). The amino acids which are directly in contact with the product at subsites +1 and +2 (D214, E217, H228, T246, R251, D259 and R394 in *Tr* Cel7A numbering) are conserved in *At, Ct, Ta, Te* and *Tr* Cel7A, and the only difference was detected in *Ma* Cel7B, which had a single conservative substitution Asp259Asn. However, there are differences at the tunnel exit region around the suggested +3 and +4 sites of *Tr* Cel7A (Divne *et al.*, 1998). According to 3D structural data these additional subsites +3 and +4 are not formed in *Ma* Cel7B, as the D338 (deleted in *Tr* Cel7A) prevents the binding of glycosyl units beyond subsite +2 (Parkkinen *et al.*, 2008). Interestingly, the 3D structural models of *At, Ct* and *Ta* Cel7A enzymes (IV) and the known *Te* Cel7A 3D structure suggest a similar arrangement, showing either an aspartic acid or an asparagine residue at this position. However, these putative structural differences do not appear to correlate with the differences detected in the product inhibition, as the characteristic feature of *Ma* Cel7B is very strong product inhibition (K_i = 6 µM), whereas the structurally similar (and possibly also lacking the subsites +3 and +4) *At, Ct, Ta, Te* Cel7A enzymes have clearly weaker cellobiose inhibition (K_i values from 39 µM up to 180 µM (Table 3 and (Tuohy *et al.*, 2002)). The *Tr* Cel7A inhibition constant (20 µM) falls between these extremes.

The activities of the Cel7 wild-type enzymes measured at 22°C on soluble, chromophoric substrates demonstrated that *Ct* Cel7A, as the most active enzyme in this study, was almost ten times more active on MULac and 19 times more active on CNPLac than *Ma* Cel7B (Table 3). The activities of the other Cel7 cellobiohydrolases fell between these two enzymes. The low activity of *Ma* Cel7B is not easily explainable from the 3D structure, as the catalytic and surrounding amino acid residues are identical in the *Tr* Cel7A complex structures (Parkkinen *et al.*, 2008). As discussed in (Parkkinen *et al.*, 2008), the lower activity might be due to slightly different binding of the substrate and product, caused by small changes throughout the 3D structure resulting in minor changes in the positions of active site residues.

Concerning microcrystalline cellulose hydrolysis, the best cellobiohydrolase at 45°C was *Ct* Cel7A. After 24h hydrolysis *Ct* Cel7A solubilised 10% of Avicel, whereas *Ma* Cel7B + *Tr* CBM solubilised only 3%, the yields from *At* Cel7A, *Ta* Cel7A+*Tr/Ct* CBM and *Tr* Cel7A being 6–8%. Factors contributing to the good performance of *Ct* Cel7A on Avicel could include an optimal active site tunnel structure and a relatively low cellobiose inhibition combined with a high-affinity CBM (see also below). However, *Ta* Cel7A+*Ct* CBM was the most effective enzyme for microcrystalline cellulose hydrolysis at elevated temperature (70°C),

even though it has similar unfolding temperature to *Ct* Cel7A. Similar result was obtained also with the catalytic modules of *Ct* Cel7A and *Ta* Cel7A, as *Ta* Cel7A had the highest hydrolysis efficiency in 70°C (unpublished results). The high activity of *Ta* Cel7A in 70°C could result from better stabilization by polymeric substrate or from differences in the product (cellobiose) inhibition on the crystalline substrate. The fusion of *Ta* Cel7A with *Tr* CBM showed slightly reduced activity in the high temperature, indicating that the thermoactivity of *Ta* Cel7A+*Ct* CBM would also be partially related to the CBM and the linker. Direct comparison of the Avicel hydrolysis efficiency of the single-module enzyme *Te* Cel7A (Publication III, Fig. 6A), can be done only with another single-module enzyme *Ma* Cel7B (Publication II, Fig. 7A), which has been measured in similar conditions using 5% substrate at 70°C. The comparison reveals that the *Te* Cel7A wt is better on Avicel than *Ma* Cel7B.

4.3 Role of CBM

Degradation of highly-ordered crystalline cellulose by the two-module *T. reesei* Cel7A enzyme has been well characterized and the important role of the CBM in the adsorption and hydrolysis has been reported (Ståhlberg *et al.*, 1991; van Tilbeurgh *et al.*, 1986). However, there are some cellobiohydrolases which do not possess CBM in nature, for example the *Ma, Ta* and *Te* Cel7 enzymes studied here. When a CBM was linked to *Ma* and *Ta* Cel7 enzymes (Publications II and IV), a clear improvement in Avicel hydrolysis was detected. The three different CBMs in this study from *At* Cel7A, *Ct* Cel7A and *Tr* Cel7A belong to CBM-1 family and have conserved sequences. They all are 36 aa in length, and stabilized by 2 or 3 S-S bridges. *Tr* Cel7A CBM interacts with cellulose through a hydrophobic flat surface formed by conserved aromatic amino acids (Linder *et al.*, 1995) called the binding surface. The aromatic amino acids in the binding face of *Tr* Cel7A CBM are Y5, Y31 and Y32 (Fig. 6), and due to the high sequence conservation the aromatic amino acids in the corresponding positions in *At* and *Ct* Cel7A CBMs (W5, W31 and Y32) are presumed to form a similar binding surface (boxed residues in the sequence alignment, Fig 8). High activity towards crystalline cellulose was reported to correlate to the high cellulose-binding affinity when *Humicola grisea* Cel7A CBM binding face was studied by mutagenesis (Takashima *et al.*, 2007). The best binding (and activity) in that case was achieved with the same pattern of aromatic amino acids as seen in *At* and *Ct* Cel7A CBMs. The binding face

similar to *Tr* Cel7A CBM was slightly less favourable (Takashima *et al.*, 2007). Similar differences between these two types of binding surfaces can be seen in the Avicel hydrolysis in this study, as the *Ta* Cel7A catalytic module fusion with *Ct* Cel7A CBM gave reproducibly slightly higher activities than the *Tr* Cel7A CBM fusion (Publication IV, Fig 4C). The differences detected between these fusion proteins could arise also from the differences in the liker region or the connection between the linker and the catalytic module.

4.4 Thermostable mutants

Although factors affecting the stability of protein structure have been well studied, the effects of individual mutations are difficult to predict. Instead of site-directed mutagenesis, the protein sequence can be subjected to random mutagenesis followed by functional screening for improved properties. Results from random mutagenesis can be used to study the relationship between the primary sequence and the properties of the folded protein. When the mutations in the nine thermostable *Ma* Cel7A mutants, discovered from the random mutant library (Publication I), were evaluated using the 3D structure, the potential role in stabilization of three individual amino acid substitutions (A30T, G184D and S290T) could be rationalized (I). Mutation G184D appeared to stabilize the structure by forming an additional salt-bridge with K186, because of the mutation the negatively charged aspartate was placed in close proximity to a positively charged lysine. Such electrostatic interactions have often been reported to be the most prominent differences between mesophilic and thermophilic protein structures (see section 1.6.2 *Structural reasons for higher thermostability* and for example (Karshikoff and Ladenstein, 2001). Substitutions A30T and S290T appear to improve packing through filling a cavity in the interior of the protein. Cavity filling has often been used as one of the strategies for site-directed mutagenesis to improve thermostability. For example, the thermostablility of chicken lysozyme was significantly increased by cavity filling mutations, making the hydrophobic core of the protein more compact (Shih and Kirsch, 1995). However, other studies (Bueno *et al.*, 2006; Eijsink *et al.*, 1992) have shown that cavity filling mutations can have more moderate effects on the thermostability than expected, suggesting that even small substitutions require some destabilising rearrangements, and that mutations can have both negative and positive effects on stability. In general, there are numerous possibilities to introduce mutations of this type into a protein structure, and it would have been difficult to design them without random mutagenesis

and functional screening. In the case of S290T mutation, sequence alignment of the GH-7 family could also have been used to identify it as a stabilising mutation, since most of the cellobiohydrolases in the GH-7 family have a threonine residue at the corresponding position. This observation supports the solidity of the consensus concept for finding stabilizing mutations; according to Lehmann (Lehmann and Wyss, 2001) stability can be improved by designing amino acid substitution for one that is predominant in the family. However, such a sequence analysis also generates many other equally good candidates.

The rational mutagenesis in this study included rigidifying mutations (S-S bridges) for thermostability. All six disulphide bridges in *Ma* Cel7B and *Te* Cel7A catalytic modules were designed on the surface of the proteins, on flexible loops near or forming the active site tunnel of the Cel7 enzymes (II, III). Flexibility of these sites was indicated by high crystallographic thermal factors (III) or by the location close to the N-terminal end of the polypeptide chain (II, III). Flexible surface loops have been found to function as initiation points for the thermal unfolding process, allowing solvent penetration in the core of the protein leading to unfolding (Colombo and Merz, 1999), which suggests that it might be possible to improve the thermostability by restricting the mobility of the loops. In addition, the placement of the S-S bridges in flexible sites allows local re-adjustment to optimal disulphide bridge geometry and minimization of the interference caused by cysteine residues (Matsumura *et al.*, 1989).

One of the individual S-S mutants in *Te* Cel7A (G4C/A72C, located near the N-terminus), which led to the greatest improvement in the unfolding temperature (ΔT_m = 5°C), was also shown to increase the thermostability of *Ma* Cel7A (ΔT_m = 2.5°C) (II, III). This disulphide bridge apparently stabilizes the structures by fixing the N-terminus of the polypeptide chain close to the protein core. In general, both the number and positions of the S-S bridges are strictly conserved in GH-7 family cellobiohydrolases, most of them containing 9, and some 10 S-S bridges. Both *Ma* Cel7B and *Te* Cel7A contain 9 S-S bonds in nature (Grassick *et al.*, 2004; Parkkinen *et al.*, 2008). The position of the 10^{th} S-S bond was adopted from *Tr* Cel7A structure. As the addition of the 10^{th} disulphide bridge clearly improved the thermostability and thermoactivity of the two cellobiohydrolases studied here, it could be a general concept to stabilize GH-7 family cellobiohydrolases lacking the 10^{th} S-S bridge.

The other S-S bridge mutations in Publication III, which improved the thermostability of *Te* Cel7A (N54C/P191C and T243C/A375C) stabilize the tunnel-forming loops 1 (residues 41–59), 3 (residues 175–205), 4 (residues 228–252)

and 6 (residues 373–398). The T_m values of these mutants were improved by 3.5°C (from 75°C to 78.5°C) and 4°C (from 75°C to 79°C), respectively, in addition to clearly prolonged half-life times ($t_{1/2}$ DTT) (III). Disulphide bridge C54-C191 connects the neighbouring loops 1 and 3 on one side of the tunnel, whereas disulphide bridge C243-C375 connects the loops 4 and 6 across the substrate binding tunnel. The restricted mobility of the loops 1 and 3 also led somewhat surprisingly to improved activity against both soluble and insoluble substrate, whereas the S-S bridge mutant T243C/A375C, in which the loops 4 and 6 were attached to each other over the substrate binding tunnel, had lower activity on crystalline cellulose than the wild-type enzyme, except in high temperature (75°C), when the wild-type enzyme had already lost most of its activity due to thermal unfolding. Interestingly, the product profile Avicel hydrolysis of T243C/A375C mutant was identical to that of the wild-type enzyme; both enzymes produced mainly cellobiose and low amounts of glucose and cellotriose. These results would indicate that the full flexibility of the loops 4 and 6 is not crucial for the mode of action of *Te* Cel7A, differing for example from bacterial cellobiohydrolases belonging to the GH-6 family (Zhang *et al.*, 2000), in which added disulphide bridges in the tunnel-forming loops caused weaker binding to ligands and lower activities and processivity. These three additional disulphide bridges were combined in a single *Te* Cel7A mutant (Publication III), resulting in a highly thermostable Cel7A which could hydrolyse microcrystalline cellulose (Avicel) at 80°C. To our knowledge this is the highest hydrolysis temperature obtained for any GH-7 cellobiohydrolases. The effects of the mutations were at least partially additive, since the triple mutant demonstrated 9°C improvement in T_m, while the individual S-S mutants improved the T_m by 3.5 to 5°C. The triple mutant had the highest T_m (84°C) reported for any GH-7 cellobiohydrolases.

5. Conclusions and future perspectives

The growing energy demand and the need for decreasing dependence on fossil fuels are great challenges. Sustainable energy production and reduction of oil dependence will require new applications, and biotechnology has been shown to have high potential. The production of bioethanol for transportation fuel is already a well established technology. However, the current systems utilising sugar cane (Brazil) and corn (US) as raw materials compete with food and feed production and are therefore not accepted as the ultimate solution for fulfilling the world's energy demand. Production of bioethanol from cellulosic raw materials would open opportunities for the use of non-food raw materials such as all kinds of wood and other biomass feedstock such as agricultural waste. Ethanol production from cellulose based materials still requires research and development and is not currently in commercial production. Cost-effective ethanol production from these raw materials would benefit from more efficient hydrolysis by novel or improved cellulases. For bioethanol production, as well as for other industrial applications in which cellulases are utilised, both improved stability and activity would be advantageous.

Altough heterologous expression of fungal Cel7 cellobiohydrolases has been difficult leading to low production levels and over-glycosylated product; here the expression of both the single-module enzymes tested, *Ma* Cel7B and *Te* Cel7A proved to be successful, thus allowing proper protein engineering studies. Two different mutagenesis strategies, random- and site-directed mutagenesis, were used to improve the thermostabilities and activities of these fungal cellobiohydrolases. Substantial improvement was achieved especially by site-directed mutagenesis; *Te* Cel7A mutant containing three additional disulphide bridges was active against crystalline cellulose substrate even at 80°C and had an unfolding temperature of 84°C, the highest reported for an enzyme belonging to GH family 7. It is probable that by fusing a fungal CBM to this *Te* Cel7 mutant (catalytic

module), one could create a fully active thermostable enzyme on crystalline substrates, similarly as shown with *Ma* Cel7B and *Ta* Cel7A fusion enzymes.

The screening system developed in this study can be used to obtain mutants that are advantageous for industrial applications. In future, random mutagenesis can be used more efficiently, since development of gene synthesis technology has enabled production of precisely specified libraries of variant genes. It become evident that the Cel7 enzymes behave differently on soluble and insoluble substrates at high temperatures, which might indicate differences in how the cellobiohydrolases are stabilized by the polymeric substrate or differences in product (cellobiose) inhibition as a function of temperature. Thus, it is of importance to use the application conditions, e.g. real substrate and temperature as early as possible in the HTS screenings.

In addition to the mutagenesis studies, new fungal cellobiohydrolases were cloned and characterized in order to identify enzymes which would be highly active on insoluble cellulose. The kinetic and stability properties of these three cellobiohydrolases were compared to those of one of the best characterised cellobiohydrolases, Cel7A from *T. reesei* and their 3D structures were studied through homology models. As evident all three Cel7 cellobiohydrolases were better than the reference enzyme in terms of both thermostability and activity on crystalline substrate, and are promising enzymes for application purposes.

References

Akasako, A., M. Haruki, M. Oobatake, and S. Kanaya. 1997. Conformational stabilities of *Escherichia coli* RNase HI variants with a series of amino acid substitutions at a cavity within the hydrophobic core. J Biol Chem 272:18686–93.

Alder, A.J., N.J. Greenfield, and G.D. Fasman. 1973. Circular dichroism and optical rotatory dispersion of proteins and polypeptides. Methods in Enzymol. 27:675–735.

Amin, N., A.D. Liu, S. Ramer, W. Aehle, D. Meijer, M. Metin, S. Wong, P. Gualfetti, and V. Schellenberger. 2004. Construction of stabilized proteins by combinatorial consensus mutagenesis. Protein Eng Des Sel 17:787–93.

Arnold, F.H., P.L. Wintrode, K. Miyazaki, and A. Gershenson. 2001. How enzymes adapt: lessons from directed evolution. TRENDS Biochem Sci 26:100–106.

Atalla, R.H., and D.L. Van der Hart. 1984. Native Cellulose: A Composite of Two Distinct Crystalline Forms. Science 223:283–285.

Banerjee, S., S. Mudliar, R. Sen, B. Giri, D. Satpute, T. Chakrabarti, and R.A. Pandey. 2010. Commercializing lignocellulosic bioethanol: technology bottlenecks and possible remedies. Biofuels, Biofuel Bioprod Bior 4:77–93.

Bauer, M.W., and R.M. Kelly. 1998. The family 1 beta-glucosidases from Pyrococcus furiosus and Agrobacterium faecalis share a common catalytic mechanism. Biochemistry 37:17170–8.

Bayer, E., Y. Shoham, and R. Lamed. 2006. Cellulose-decomposing Bacteria and Their Enzyme Systems Springer, New York.

Bayer, E.A., J.P. Belaich, Y. Shoham, and R. Lamed. 2004. The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. Annu Rev Microbiol. 58:521–54.

Bayer, E.A., H. Chancy, R. Lamed, and Y. Sholam. 1998. Cellulose, cellulases and cellulosomes. Curr Opin Struct Biol 8:548–557.

Becker, D., C. Braet, H. Brumer, 3rd, M. Claeyssens, C. Divne, B.R. Fagerstrom, M. Harris, T.A. Jones, G.J. Kleywegt, A. Koivula, S. Mahdi, K. Piens, M.L. Sinnott, J. Stahlberg, T.T. Teeri, M. Underwood, and G. Wohlfahrt. 2001. Engineering of a glycosidase Family 7 cellobiohydrolase to more alkaline pH optimum: the pH behaviour of *Trichoderma reesei* Cel7A and its E223S/ A224H/L225V/T226A/D262G mutant. Biochem J 356:19–30.

Bhat, M.K. 2000. Cellulases and related enzymes in biotechnology. Biotech Adv 18:355–383.

Boer, H., T.T. Teeri, and A. Koivula. 2000. Characterization of *Trichoderma reesei* cellobiohydrolase Cel7A secreted from *Pichia pastoris* using two different promoters. Biotechnol Bioeng 69:486–494.

Boisset, C., C. Fraschini, M. Schulein, B. Henrissat, and H. Chanzy. 2000. Imaging the enzymatic digestion of bacterial cellulose ribbons reveals the endo character of the cellobiohydrolase Cel6A from *Humicola insolens* and its mode of synergy with cellobiohydrolase Cel7A. Appl Environ Microbiol 66:1444–52.

Boraston, A.B., D.N. Bolam, H.J. Gilbert, and G.J. Davies. 2004. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem J 382:769–781.

Brannigan, J.A., and A.J. Wilkinson. 2002. Protein engineering 20 years on. Nat Rev Mol Cell Biol 3:964–70.

Bueno, M., N. Cremades, J.L. Neira, and J. Sancho. 2006. Filling small, empty protein cavities: structural and energetic consequences. J Mol Biol 358:701–12.

Bulter, T., M. Alcalde, V. Sieber, P. Meinhold, C. Schlachtbauer, and F.H. Arnold. 2003. Functional expression of a fungal laccase in *Saccharomyces cerevisiae* by directed evolution. Appl Environ Microb 69:987–995.

Cherry, J.R., M.H. Lamsa, P. Schneider, J. Vind, A. Svendsen, A. Jones, and A.H. Pedersen. 1999. Directed evolution of fungal peroxidase. Nat Biotechnol 17:379–384.

Chirakkal, H., G.C. Ford, and A. Moir. 2001. Analysis of a conserved hydrophobic pocket important for the thermostability of Bacillus pumilus chloramphenicol acetyltransferase (CAT-86). Protein Eng 14:161–6.

Claeyssens, M., H. Van Tilbeurgh, P. Tomme, T.M. Wood, and S.I. McRae. 1989. Fungal cellulase systems. Comparison of the specificities of the cellobiohydrolases isolated from *Penicillium pinophilum* and *Trichoderma reesei*. Biochem J. 261:819–25.

Coco, W.M., W.E. Levinson, M.J. Crist, H.J. Hektor, A. Darzins, P.T. Pienkos, C.H. Squires, and D.J. Monticello. 2001. DNA shuffling method for generating highly reconbined genes and evolved enzymes. Nat Biotechnol 19:354–359.

Colombo, C., and K.M. Merz. 1999. Stability and Activity of Mesophilic Subtilisin E and Its Thermophilic Homolog: Insights from Molecular Dynamics Simulations. J Am Chem Soc 121:6895–6903.

Crameri, A., S.A. Raillard, E. Bermudez, and W.P.C. Stemmer. 1998. DNA shuffling of a family of genes from diverse species accelerates directed evolution. Nature 391:288–291.

Cullen, D., and P.J. Kersten. 2004. Enzymology and Molecular Biology of Lignin Degradation, In: R. B. a. G. A. M. (Eds.) The Mycota III, Biochemistry and Molecular Biology, 2nd ed. Springer-Verlag, Berlin-Heidelberg.

Daggett, V., and M. Levitt. 1992. A model of the molten globule state from molecular dynamics simulations. Proc Natl Acad Sci U S A 89:5142–6.

Dalby, P.A. 2003. Optimising enzyme function by directed evolution. Curr Opin Struc Biol 13:1–6.

Das, R., and M. Gerstein. 2000. The stability of thermophilic proteins: a study based on comprehensive genome comparison. Funct Integr Genomics 1:76–88.

Davies, G., and B. Henrissat. 1995. Structures and mechanisms of glycosyl hydrolases. Structure 3:853–9.

Davies, G.J., A.M. Brzozowski, M. Dauter, A. Varrot, and M. Schulein. 2000. Structure and function of *Humicola insolens* family 6 cellulases: structure of the endoglucanase, Cel6B, at 1.6 A resolution. Biochem J 348 Pt 1:201–7.

Davies, G.J., V. Ducros, R.J. Lewis, T.V. Borchert, and M. Schulein. 1997. Oligosaccharide specificity of a family 7 endoglucanase: insertion of potential sugar-binding subsites. J Biotechnol 57:91–100.

Davies, G.J., S.J. Gamblin, J.A. Littlechild, and H.C. Watson. 1993. The structure of a thermally stable 3-phosphoglycerate kinase and a comparison with its mesophilic equivalent. Proteins 15:283–9.

Dawson, R.M.C., E.D. C., W.H. Elliot, and K.M. Jones. 1959. Data for biochemical research. Oxford University Press, Amen House, London.

Dighe, A.S., P.M. Patel, and K.K. Rao. 1985. Effect of phosphate on the estimation of reducing sugars in biological media. Biotechnol Bioeng 27:1612–1615.

Dill, K.A. 1990. Dominant forces in protein folding. Biochemistry. 29:7133–55.

Ding, S.Y., and M.E. Himmel. 2006. The maize primary cell wall microfibril: a new model derived from direct visualization. J Agric Food Chem 54:597–606.

Divne, C., J. Ståhlberg, T.T. Teeri, and T.A. Jones. 1998. High-resolution crystal structures reveal how a cellulose chain is bound in the 50 Å long tunnel of cellobiohydrolase I from *Trichoderma reesei*. J. Mol. Biol. 275:309–325.

Divne, C., J. Ståhlberg, T. Reinikainen, L. Ruohonen, G. Pettersson, J.K. Knowles, T.T. Teeri, and T.A. Jones. 1994. The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. Science 265:524–528.

Duff, S., and W. Murray. 1996. Bioconversion of forest products industry waste cellulosics to fuel ethanol: a review. Bioresource Technol. 55:1–33.

Eggert, T., S.A. Funke, N.M. Rao, P. Acharya, H. Krumm, M.T. Reetz, and K.-E. Jaeger. 2005. Multiplex-PCR-based recombination as a novel high-fidelity method for directed evolution. ChemBioChem 6:1062–1067.

Eijsink, V.G.H., A. Bjørk, S. Gåseidnes, R. Sirevåg, B. Synstad, B. van den Burg, and G. Vriend. 2004. Rational engineering of enzyme stability. J Biotechnol 113:105–120.

Eijsink, V.G., B.W. Dijkstra, G. Vriend, J.R. van der Zee, O.R. Veltman, B. van der Vinne, B. van den Burg, S. Kempe, and G. Venema. 1992. The effect of cavity-filling mutations on the thermostability of Bacillus stearothermophilus neutral protease. Protein Eng 5:421–6.

Erbel, P.J., Y. Karimi-Nejad, T. De Beer, R. Boelens, J.P. Kamerling, and J.F. Vliegenthart. 1999. Solution structure of the alpha-subunit of human chorionic gonadotropin. Eur J Biochem 260:490–8.

Ericsson, K., and L.J. Nilsson. 2006. Assessment of the potential biomass supply in Europe using a resource-focused approach. Biomass Bioeng 30:1–15.

Eriksson, K.E.L., R.A. Blanchette, and P. Ander. 1990. Microbial and enzymatic degradation of wood and wood components Springer-Verlag, New York, N.Y.

Foreman, P.K., D. Brown, L. Dankmeyer, R. Dean, S. Diener, N.S. Dunn-Coleman, F. Goedegebuur, T.D. Houfek, G.J. England, A.S. Kelley, H.J. Meerman, T. Mitchell, C. Mitchinson, H.A. Olivares, P.J. Teunissen, J. Yao, and M. Ward. 2003. Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma reesei*. J Biol Chem 278:31988–97.

Forouchi, E., and D.J. Gunn. 1983. Some effects of metal ions on the estimation of reducing sugars in biological media. Biotechnol Bioeng 25:1905–1911.

Freer, S.N. 1993. Kinetic characterization of a beta-glucosidase from a yeast, *Candida wickerhamii*. J Biol Chem 268:9337–42.

Ghose, T.K. 1987. Measurement of cellulase activities. Pure & Appl. Chem. 59:257–268.

Gilkes, N.R., B. Henrissat, D.C. Kilburn, R.C. Miller, Jr., and R.A.J. Warren. 1991. Domains in microbial beta-1,4-glycanases: sequence conservation, function, and enzyme families. Microbiol Rev 55:303–315.

Giver, L., A. Gershenson, P.O. Freskgard, and F.H. Arnold. 1998. Directed evolution of a thermostable esterase. Proc Natl Acad Sci USA 95:12809–12813.

Gnansounou, E. 2010. Production and use of lignocellulosic bioethanol in Europe: Current situation and perspectives. Bioresour Technol 101:4842–50.

Goodenough, P.W., and J.A. Jenkins. 1991. Protein engineering to change thermal stability for food enzymes. Biochem Soc Trans 19:655–62.

Grassick, A., P.G. Murray, R. Thompson, C.M. Collins, L. Byrnes, G. Birrane, T.M. Higgins, and M.G. Tuohy. 2004. Three-dimensional structure of a thermostable native cellobiohydrolase, CBH IB, and molecular characterization of the cel7 gene from the filamentous fungus, *Talaromyces emersonii*. Eur J Biochem. 271:4495–4506.

Guex, N., and M.C. Peitsch. 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18:2714–2723.

Haakana, H., A. Miettinen-Oinonen, V. Joutsjoki, A. Mäntylä, P. Suominen, and J. Vehmaanperä. 2004. Cloning of cellulase genes from *Melanocarpus albomyces* and their efficient expression in *Trichoderma reesei*. Enzyme Microb. Technol. 34:159–167.

Haney, P.J., J.H. Badger, G.L. Buldak, C.I. Reich, C.R. Woese, and G.J. Olsen. 1999. Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic *Methanococcus* species. Proc Natl Acad Sci U S A 96:3578–83.

Harayama, S. 1998. Artificial evolution by DNA shuffling. Tibtech 16:76–82.

Harjunpää, V., A. Teleman, A. Koivula, L. Ruohonen, T.T. Teeri, O. Teleman, and T. Drakenberg. 1996. Cello-oligosaccharide hydrolysis by cellobiohydrolase II from *Trichoderma reesei*. Association and rate constants derived from an analysis of progress curves. Eur J Biochem 240:584–91.

Harris, E.M., A.E. Aleshin, L.M. Firsov, and R.B. Honzatko. 1993. Refined structure for the complex of 1-deoxynojirimycin with glucoamylase from Aspergillus awamori var. X100 to 2.4-A resolution. Biochemistry 32:1618–26.

Harrison, M.J., A.S. Nouwens, D.R. Jardine, N.E. Zachara, A.A. Gooley, H. Nevalainen, and N.H. Packer. 1998. Modified glycosylation of cellobiohydrolase I from a high cellulase-producing mutant strain of *Trichoderma reesei*. Eur J Biochem 256:119–27.

Hayashi, N., J. Sugiyama, T. Okano, and M. Ishihara. 1997. Selective degradation of the cellulose I(a) component in *Cladophora* cellulose with *Trichoderma viride* cellulase. Carbohyd Res 305:109–116.

Hehre, E.J. 2000. A fresh understanding of the stereochemical behaviour of glycosylases: structural distinction of "inverting" (2-MCO-type) versus "retaining" (1-MCO-type) enzymes, *In* D. Horton, ed. Advances in Carbohydrate Chemistry and Biochemistry, Vol. 55. Academic press, San Diego, California.

Heinzelman, P., C.D. Snow, I. Wu, C. Nguyen, A. Villalobos, S. Govindarajan, J. Minshull, and F.H. Arnold. 2009. A family of thermostable fungal cellulases created by structure-guided recombination. Proc Natl Acad Sci U S A 106:5610–5.

Henrissat, B., and A. Bairoch. 1996. Updating the sequence-based classification of glycosyl hydrolases. Biochem J 316:695–696.

Henrissat, B., T.T. Teeri, and R.A. Warren. 1998. A scheme for designating enzymes that hydrolyse the polysaccharides in the cell walls of plants. FEBS Lett 425:352–354.

Henrissat, B., I. Callebaut, S. Fabrega, P. Lehn, J.P. Mornon, and G. Davies. 1995. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. Proc Natl Acad Sci USA 92:7090–7094.

Higgins, D.G., J.D. Thompson, and T.J. Gibson. 1996. Using CLUSTAL for multiple sequence alignments. Methods Enzymol. 266:383–402.

Hon, D.N.-S. 1994. Cellulose: a random walk along its historical path. Cellulose 1:1–25.

Hoshino, E., M. Shiroishi, Y. Amano, M. Nomura, and T. Kanda. 1997. Synergistic actions of exo-type cellulases in the hydrolysis of cellulose with different crystallinities. J. Ferment. Bioeng. 84:300–306.

Hughes, S.R., S.B. Riedmuller, J.A. Mertens, X.-L. Li, K.M. Bischoff, N. Qureshi, M.A. Cotta, and P.J. Farrelly. 2006. High-throughput screening of cellulase F mutants from multiplexed plasmid sets using an automated plate assay on a functional proteomic robotic workcell. Proteome Science 4:Doi:10.1186/1477-5956-4-10.

Hulme, M.A. 1988. Viscosimetric determination of carboxymethylcellulase activity. Methods Enzymol 160:130–135.

Igarashi, K., A. Koivula, M. Wada, S. Kimura, M. Penttilä, and M. Samejima. 2009. High speed atomic force microscopy visualizes processive movement of *Trichoderma reesei* cellobiohydrolase I on crystalline cellulose. J Biol Chem 284:36186–90.

Igarashi, K., M. Wada, R. Hori, and M. Samejima. 2006. Surface density of cellobiohydrolase on crystalline celluloses. A critical parameter to evaluate enzymatic kinetics at a solid-liquid interface. Febs J 273:2869–78.

Imberty, A., and S. Perez. 1995. Stereochemistry of the N-glycosylation sites in glycoproteins. Protein Eng 8:699–709.

Imperiali, B., and S.E. O,'Connor. 1999. Effect of N-linked glycosylation on glycopeptide and glycoprotein structure. Curr Opin Chem Biol 3:643–9.

Irwin, D.C., M. Spezio, L.P. Walker, and D.B. Wilson. 1993. Activity Studies of 8 Purified Cellulases – Specificity, Synergism, and Binding Domain Effects. Biotech Bioeng 42:1002–1013.

Jaenicke, R., H. Schurig, N. Beaucamp, and R. Ostendorp. 1996. Structure and stability of hyperstable proteins: glycolytic enzymes from hyperthermophilic bacterium *Thermotoga maritima*. Adv Protein Chem 48:181–269.

Jaenicke, R., and G. Bohm. 1998. The stability of proteins in extreme environments. Curr Opin Struct Biol 8:738–48.

Jeoh, T., W. Michener, M.E. Himmel, S.R. Decker, and W.S. Adney. 2008. Implications of cellobiohydrolase glycosylation for use in biomass conversion. Biotechnol Biofuels 1:10.

Joern, J.M., P. Meinhold, and F.H. Arnold. 2002. Analysis of shuffled gene libraries. J. Mol. Biol. 316:643–656.

Johnston, D.B., S.P. Shoemaker, G.M. Smith, and J.R. Whitaker. 2007. Kinetic measurements of cellulase activity on insoluble substrates using disodium 2,2' bicinchoninate. Journal of Food Biochemistry 22:301–319.

Joo, H., A. Arisawa, Z. Lin, and F.H. Arnold. 1999. A high-throughput digital imaging screen for the discovery and directed evolution of oxygenases. Chem Biol 6:699–706.

Jänis, J., J. Rouvinen, M. Leisola, O. Turunen, and P. Vainiotalo. 2001. Thermostability of endo-1,4-beta-xylanase II from *Trichoderma reesei* studied by electrospray ionization Fourier-transform ion cyclotron resonance MS, hydrogen/deuterium-exchange reactions and dynamic light scattering. Biochem J 356:453–460.

Kanaya, S., and M. Itaya. 1992. Expression, purification, and characterization of a recombinant ribonuclease H from *Thermus thermophilus* HB8. J Biol Chem 267:10184–92.

Karhunen, T., A. Mäntylä, K.M. Nevalainen, and P.L. Suominen. 1993. High frequency one-step gene replacement in *Trichoderma reesei*. I. Endoglucanase I overproduction. Mol Gen Genet 241:515–522.

Karshikoff, A., and R. Ladenstein. 2001. Ion pairs and the thermotolerance of proteins from hyperthermophiles: a "traffic rule" for hot roads. Trends Biochem Sci 26:550–6.

Karshikoff, A., and R. Ladenstein. 1998. Proteins from thermophilic and mesophilic organisms essentially do not differ in packing. Protein Eng 11:867–72.

Kataeva, I.A., R.D. Seidel, 3rd, A. Shah, L.T. West, X.L. Li, and L.G. Ljungdahl. 2002. The fibronectin type 3-like repeat from the *Clostridium thermocellum* cellobiohydrolase CbhA promotes hydrolysis of cellulose by modifying its surface. Appl Environ Microbiol 68:4292–300.

Kidby, D.K., and D.J. Davidson. 1973. A convenient ferricyanide estimation of reducing sugars in the nanomole range. Anal Biochem. 55:321–325.

Kikuchi, M., K. Ohnishi, and S. Harayama. 2000. An effective family shuffling method using single-stranded DNA. Gene 243:133–137.

Kim, T.H., and Y.Y. Lee. 2007. Pretreatment of corn stover by soaking in aqueous ammonia at moderate temperatures. Appl Biochem Biotechnol 137–140:81–92.

Kim, Y.W., J.H. Choi, J.W. Kim, C. Park, H. Cha, S.B. Lee, B.H. Oh, T.W. Moon, and K.H. Park. 2003. Directed evolution of *Thermus maltogenic* amylase toward enhanced thermal resistance. Appl Environ Microbiol 69:4866–4874.

Kirk, O., T.V. Borchert, and C.C. Fuglsang. 2002. Industrial enzyme applications. Curr Opin Biotechnol 13:345–51.

Kleman-Leyer, K.M., M. Siika-Aho, T.T. Teeri, and T.K. Kirk. 1996. The Cellulases Endoglucanase I and Cellobiohydrolase II of *Trichoderma reesei* Act Synergistically To Solubilize Native Cotton Cellulose but Not To Decrease Its Molecular Size. Appl Environ Microbiol 62:2883–2887.

Klemm, D., B. Heublein, H.P. Fink, and A. Bohn. 2005. Cellulose: fascinating biopolymer and sustainable raw material. Angew Chem Int Ed Engl 44:3358–93.

Kleywegt, G.J., J.Y. Zou, C. Divne, G.J. Davies, I. Sinning, J. Stahlberg, T. Reinikainen, M. Srisodsuk, T.T. Teeri, and T.A. Jones. 1997. The crystal structure of the catalytic core domain of endoglucanase I from *Trichoderma reesei* at 3.6 Å resolution, and a comparison with related enzymes. J Mol Biol 272:383–97.

Knapp, S., W.M. de Vos, D. Rice, and R. Ladenstein. 1997. Crystal structure of glutamate dehydrogenase from the hyperthermophilic eubacterium *Thermotoga maritima* at 3.0 Å resolution. J Mol Biol 267:916–32.

Koivula, A., L. Ruohonen, G. Wohlfahrt, T. Reinikainen, T.T. Teeri, K. Piens, M. Claeyssens, M. Weber, A. Vasella, D. Becker, M.L. Sinnott, J.Y. Zou, G.J. Kleywegt, M. Szardenings, J. Ståhlberg, and T.A. Jones. 2002. The active site of cellobiohydrolase Cel6A from *Trichoderma reesei*: the roles of aspartic acids D221 and D175. J Am Chem Soc 124:10015–24.

Koivula, A., T. Reinikainen, L. Ruohonen, A. Valkeajarvi, M. Claeyssens, O. Teleman, G.J. Kleywegt, M. Szardenings, J. Rouvinen, T.A. Jones, and T.T. Teeri. 1996. The active site of *Trichoderma reesei* cellobiohydrolase II: the role of tyrosine 169. Protein Eng 9:691–9.

Koshland, D.E.J. 1953. Stereochemistry and the mechanism of enzymatic reactions. Biological Reviews 28:416–436.

Kraulis, J., G.M. Clore, M. Nilges, T.A. Jones, G. Pettersson, J. Knowles, and A.M. Gronenborn. 1989. Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei*. A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. Biochemistry 28:7241–7257.

Kubicek, C.P. 1992. The cellulase proteins of *Trichoderma reesei*: Structure, multiplicity, mode of action and regulation of formation. Adv Biochem Eng 45:1–27.

Kumar, S., and R. Nussinov. 2001. How do thermophilic proteins deal with heat? Cell Mol Life Sci 58:1216–1233.

Kumar, S., C.J. Tsai, and R. Nussinov. 2000. Factors enhancing protein thermostability. Protein Eng 13:179–191.

Kunst, A., B. Draeger, and J. Ziegenhorn. 1984. Colorimetric methods with glucose oxidase and peroxidase, p. 178–185, *In* H. U. Bergmeyer, et al., eds. Methods of enzymatic analysis, Vol. 6. Verlag Chemie, Deerfield Beach.

Lamed, R., J. Naimark, E. Morgenstern, and E.A. Bayer. 1987. Specialized cell surface structures in cellulolytic bacteria. J Bacteriol 169:3792–800.

Laymon, R.A., W.S. Adney, A. Mohagheghi, M.E. Himmel, and S.R. Thomas. 1996. Cloning and expression of full-length *Trichoderma reesei* cellobiohydrolase I cDNAs in *Escherichia coli*. Appl Biochem Biotechnol 57–58:389–97.

Lazaridis, T., I. Lee, and M. Karplus. 1997. Dynamics and unfolding pathways of a hyperthermophilic and a mesophilic rubredoxin. Protein Sci 6:2589–605.

Lehmann, M., and M. Wyss. 2001. Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. Curr Opin Biotechnol 12:371–375.

Lehmann, M., C. Loch, A. Middendorf, D. Studer, S.F. Lassen, L. Pasamontes, A.P.G.M. van Loon, and M. Wyss. 2002. The consensus concept for thermostability engineering of proteins: further proof of concept. Protein Eng 15:403–411.

Lehmann, M., L. Pasamontes, S.F. Lassen, and M. Wyss. 2000a. The consensus concept for thermostability engineering of proteins. Biochim Biophys Acta 1543:408–415.

Lehmann, M., D. Kostrewa, M. Wyss, R. Brugger, A. D'Arcy, L. Pasamontes, and A.P. van Loon. 2000b. From DNA sequence to improved functionality: using protein sequence comparisons to rapidly design a thermostable consensus phytase. Protein Eng 13:49–57.

Leung, D.W., E. Chen, and D.V. Goeddel. 1989. A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. Technique 1:11–15.

Lever, M. 1972. A new reaction for colorimetric determination of carbohydrates. Anal Biochem 47:276–279.

Li, W.F., X.X. Zhou, and P. Lu. 2005. Structural features of thermozymes. Biotechnol Adv 23:271–81.

Linder, M., I. Salovuori, L. Ruohonen, and T.T. Teeri. 1996. Characterization of a double cellulose-binding domain. Synergistic high affinity binding to crystalline cellulose. J Biol Chem 271:21268–72.

Linder, M., M.L. Mattinen, M. Kontteli, G. Lindeberg, J. Stahlberg, T. Drakenberg, T. Reinikainen, G. Pettersson, and A. Annila. 1995. Identification of functionally important amino acids in the cellulose-binding domain of *Trichoderma reesei* cellobiohydrolase I. Protein Sci 4:1056–1064.

Lynd, L.R., W.H. van Zyl, J.E. McBride, and M. Laser. 2005. Consolidated bioprocessing of cellulosic biomass: an update. Curr Opin Biotechnol 16:577–83.

Lynd, L.R., P.J. Weimer, W.H. van Zyl, and I.S. Pretorius. 2002. Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol Biol Rev 66:506–577.

Maheshwari, R., G. Bharadwaj, and M.K. Bhat. 2000. Thermophilic Fungi: Their Physiology and Enzymes. Microbiol Mol Biol Rev 64:461–488.

Mais, U., A.R. Esteghlalian, J.N. Saddler, and S.D. Mansfield. 2002. Enhancing the enzymatic hydrolysis of cellulosic materials using simultaneous ball milling. Appl Biochem Biotechnol 98–100:815–32.

Mansfeld, J., G. Vriend, B.W. Dijkstra, O.R. Veltman, B. Van den Burg, G. Venema, R. Ulbrich-Hofmann, and V.G. Eijsink. 1997. Extreme stabilization of a thermolysin-like protease by an engineered disulfide bond. J Biol Chem 272:11152–11156.

Martinez, D., R.M. Berka, B. Henrissat, M. Saloheimo, M. Arvas, S.E. Baker, J. Chapman, O. Chertkov, P.M. Coutinho, D. Cullen, E.G. Danchin, I.V. Grigoriev, P. Harris, M. Jackson, C.P. Kubicek, C.S. Han, I. Ho, L.F. Larrondo, A.L. de Leon, J.K. Magnuson, S. Merino, M. Misra, B. Nelson, N. Putnam, B. Robbertse, A.A. Salamov, M. Schmoll, A. Terry, N. Thayer, A. Westerholm-Parvinen, C.L. Schoch, J. Yao, R. Barabote, M.A. Nelson, C. Detter, D. Bruce, C.R. Kuske, G. Xie, P. Richardson, D.S. Rokhsar, S.M. Lucas, E.M. Rubin, N. Dunn-Coleman, M. Ward, and T.S. Brettin. 2008. Genome sequencing and analysis of the biomass-degrading fungus Trichoderma reesei (syn. Hypocrea jecorina). Nat Biotechnol 26:553–60.

Martinez, D., L.F. Larrondo, N. Putnam, M.D. Gelpke, K. Huang, J. Chapman, K.G. Helfenbein, P. Ramaiya, J.C. Detter, F. Larimer, P.M. Coutinho, B. Henrissat, R. Berka, D. Cullen, and D. Rokhsar. 2004. Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporiu*m strain RP78. Nat Biotechnol 22:695–700.

Matsumura, M., W.J. Becktel, M. Levitt, and B.W. Matthews. 1989. Stabilization of phage T4 lysozyme by engineered disulfide bonds. Proc Natl Acad Sci U S A 86:6562–6.

Matthews, B.W., H. Nicholson, and W.J. Becktel. 1987. Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. Proc Natl Acad Sci U S A 84:6663–6667.

McCarter, J.D., and S.G. Withers. 1994. Mechanisms of enzymatic glycoside hydrolysis. Curr Opin Struct Biol 4:885–92.

McHale, A., and M.P. Coughlan. 1981. The cellulolytic system of *Talaromyces emersonii*. Purification and characterization of the extracellular and intracellulasr β-glucosidases. Biochim Biophys Acta 662:152–159.

Medve, J., J. Karlsson, D. Lee, and F. Tjerneld. 1998. Hydrolysis of microcrystalline cellulose by cellobiohydrolase I and endoglucanase II from *Trichoderma reesei*: Adsorption, sugar production pattern, and synergism of the enzymes. Biotechnol Bioeng 59:621–634.

Medve, J., J. Ståhlberg, and F. Tjerneld. 1994. Adsorption and synergism of cellobiohydrolase I and cellobiohydrolase II of Trichoderma reesei during hydrolysis of microcrystalline cellulose. Biotechnol Bioeng 44:1064–1073.

Meinke, A., N.R. Gilkes, E. Kwan, D.G. Kilburn, R.A. Warren, and R.C. Miller, Jr. 1994. Cellobiohydrolase A (CbhA) from the cellulolytic bacterium Cellulomonas fimi is a beta-1,4-exocellobiohydrolase analogous to Trichoderma reesei CBH II. Mol Microbiol 12:413–22.

Meinke, A., N.R. Gilkes, D.G. Kilburn, R.C. Miller, Jr., and R.A. Warren. 1993. Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase D (CenD), a family A beta-1,4-glucanase. J Bacteriol 175:1910–8.

Menendez-Arias, L., and P. Argos. 1989. Engineering protein thermal stability. Sequence statistics point to residue substitutions in alpha-helices. J Mol Biol 206:397–406.

Miettinen-Oinonen, A., J. Londesborough, V. Joutsjoki, R. Lantto, and J. Vehmaanperä. 2004. Three cellulases from *Melanocarpus albomyces* with applications in textile industry. Enzyme Microb Tech 34:332–341.

Miller, G.L. 1959. Use of dinitrosalicylic acid reagent for determination of reducing sugar. Anal Chem 31:426–428.

Miyazaki, K., P.L. Wintrode, R.A. Grayling, D.N. Rubingh, and F.H. Arnold. 2000. Directed evolution study of temperature adaption in a psychrophilic enzyme. J Mol Biol 297:1015–1026.

Miyazaki, K., and F.H. Arnold. 1999. Exploring Nonnatural Evolutionary Pathways by Saturation Mutagenesis: Rabid Improvement of Protein Function. J Mol Evol 49:716–720.

Moloney, A.P., S.I. McCrae, T.M. Wood, and M.P. Coughlan. 1985. Isolation and characterization of the endoglucanases of *Talaromyces emersonii*. Biochem. J. 225:365–374.

Moore, J.C., H.-M. Jin, O. Kuchner, and F.H. Arnold. 1997. Strategies for the *in vitro* evolution of protein function: enzyme evolution by random recombination of improved sequences. J. Mol. Biol. 272:336–347.

Morawski, B., Z. Lin, P. Cirino, H. Joo, G. Bandara, and F.H. Arnold. 2000. Functional expression of horseradish peroxidase in *Saccharomyces cerevisiae* and *Pichia pastoris*. Protein Eng 13:377–384.

Munoz, I.G., W. Ubhayasekera, H. Henriksson, I. Szabo, G. Pettersson, G. Johansson, S.L. Mowbray, and J. Ståhlberg. 2001. Family 7 cellobiohydrolases from *Phanerochaete chrysosporium*: crystal structure of the catalytic module of Cel7D (CBH58) at 1.32 A resolution and homology models of the isozymes. J Mol Biol. 314:1097–1111.

Murashima, K., A. Kosugi, and R.H. Doi. 2002. Thermostabilization of cellulosomal endoglucanase EngB from *Clostridium cellulovorans* by in vitro DNA recombination with non-cellulosomal endoglucanase EngD. Mol Microbiol 45:617–626.

Nelson, N. 1944. A photometric adaptation of the Somogyi method for the determination of glucose. J Biol Chem 153:375–380.

Nidetzky, B., W. Steiner, M. Hayn, and M. Claeyssens. 1994a. Cellulose hydrolysis by the cellulases from *Trichoderma reesei*: a new model for synergistic interaction. Biochem J 298 Pt 3:705–10.

Nidetzky, B., W. Zachariae, G. Gercken, M. Hayn, and W. Steiner. 1994b. Hydrolysis of cellooligosaccharides by *Trichoderma reesei* cellobiohydrolases – Experimental data and kinetic modeling. Enzyme Microb Tech 16:43–52.

Nijikken, Y., T. Tsukada, K. Igarashi, M. Samejima, T. Wakagi, H. Shoun, and S. Fushinobu. 2007. Crystal structure of intracellular family 1 beta-glucosidase BGL1A from the basidiomycete *Phanerochaete chrysosporium*. FEBS Lett 581:1514–20.

Nishiyama, Y., J. Sugiyama, H. Chanzy, and P. Langan. 2003. Crystal structure and hydrogen bonding system in cellulose I(alpha) from synchrotron X-ray and neutron fiber diffraction. J Am Chem Soc 125:14300–6.

Nishiyama, Y., P. Langan, and H. Chanzy. 2002. Crystal structure and hydrogen-bonding system in cellulose Ibeta from synchrotron X-ray and neutron fiber diffraction. J Am Chem Soc 124:9074–82.

Okada, H., K. Tada, T. Sekiya, K. Yokoyama, A. Takahashi, H. Tohda, H. Kumagai, and Y. Morikawa. 1998. Molecular characterization and heterologous expression of the gene encoding a low-molecular-mass endoglucanase from *Trichoderma reesei* QM9414. Appl Environ Microbiol 64:555–63.

Olson, L., and P. Stanley. 1990. Compositions and methods to vary color density Patent WO 90/02790 1990.

Orr-Weaver, T.L., J.W. Szostak, and R.J. Rothstein. 1981. Yeast transformation-a model system for the study of recombination. Proc Natl Acad Sci 78:6354–6358.

Ostermeier, M., J.H. Shim, and S.J. Benkovic. 1999. A combinatorial approach to hybrid enzymes independent of DNA homology. Nat Biotechnol 17:1205–1209.

Otten, L.G., C.F. Sio, J. Vrielink, R.H. Cool, and W.J. Quax. 2002. Altering the substrate specificity of cephalosporin acylase by directed evolution of the beta-subunit. J. Biol. Chem. 277:42121–42127.

Pace, C.N. 1990. Measuring and increasing protein stability. Trends Biotechnol 8:93–8.

Pace, C.N., F. Vajdos, L. Fe, G. Grimsley, and T. Gray. 1995. How to measure and predict the molar absorption coefficient of a protein. Protein Sci 4:2411–2423.

Pagan, M., R.J. Sola, and K. Griebenow. 2009. On the role of protein structural dynamics in the catalytic activity and thermostability of serine protease subtilisin Carlsberg. Biotechnol Bioeng 103:77–84.

Paloheimo, M., A. Mäntylä, J. Kallio, and P. Suominen. 2003. High-yield production of a bacterial xylanase in the filamentous fungus *Trichoderma reesei* requires a carrier polypeptide with an intact domain structure. Appl Environ Microbiol 69:7073–7082.

Palonen, H., A.B. Thomsen, M. Tenkanen, A.S. Schmidt, and L. Viikari. 2004. Evaluation of wet oxidation pretreatment for enzymatic hydrolysis of softwood. Appl Biochem Biotechnol 117:1–17.

Palonen, H., M. Tenkanen, and M. Linder. 1999. Dynamic interaction of *Trichoderma reesei* cellobiohydrolases Cel6A and Cel7A and cellulose at equilibrium and during hydrolysis. Appl Environ Microbiol 65:5229–33.

Parkkinen, T., A. Koivula, J. Vehmaanperä, and J. Rouvinen. 2008. Crystal structures of *Melanocarpus albomyces* cellobiohydrolase Cel7B in complex with cello-oligomers show high flexibility in the substrate binding. Protein Sci 17:1383–1394.

Penttilä, M.E., L. André, P. Lehtovaara, M.J. Bailey, T.T. Teeri, and J.K.C. Knowles. 1988. Efficient secretion of two fungal cellobiohydrolases by *Saccharomyces cerevisiae*. Gene 63:103–112.

Penttilä, M., H. Nevalainen, M. Rättö, E. Salminen, and J. Knowles. 1987. A versatile transformation system for the cellulolytic filamentous fungus *Trichoderma reesei*. Gene 61:155–164.

Penttilä, M., P. Lehtovaara, H. Nevalainen, R. Bhikhabhai, and J. Knowles. 1986. Homology between cellulase genes of *Trichoderma reesei*: complete nucleotide sequence of the endoglucanase I gene. Gene 45:253–63.

Pere, J., M. Siika-aho, and L. Viikari. 2000. Biomechanical pulping with enzymes: response of coarse mechanical pulp to enzymatic modification and secondary refining. TAPPI J. 83:85–185.

Petrescu, A.J., A.L. Milac, S.M. Petrescu, R.A. Dwek, and M.R. Wormald. 2004. Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. Glycobiology 14:103–14.

Pompon, D., and A. Nicolas. 1989. Protein engineering by cDNA recombination in yeasts: shuffling of mammalian cytochrome P-450 functions. Gene. 83:15–24.

Prakash, O., and N. Jaiswal. 2010. α-Amylase: an ideal representative of thermostable enzymes. Appl Biochem Biotechnol 160:2401–14.

Reinikainen, T., O. Teleman, and T.T. Teeri. 1995. Effects of pH and high ionic strength on the adsorption and activity of native and mutated cellobiohydrolase I from *Trichoderma reesei*. Proteins 22:392–403.

Reinikainen, T., L. Ruohonen, T. Nevanen, L. Laaksonen, P. Kraulis, J. T.A., J.K.C. Knowles, and T.T. Teeri. 1992. Investigation of the function of mutated cellulose-binding domains of *Trichoderma reesei* cellobiohydrolase I. Proteins: Structure, function and genetics 14:475–482.

Ross, P., R. Mayer, and M. Benziman. 1991. Cellulose biosynthesis and function in bacteria. Microbiol Rev. 55:35–58.

Rouvinen, J., T. Bergfors, T. Teeri, J.K. Knowles, and T.A. Jones. 1990. Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. Science 249:380–386.

Russell, R.J., J.M. Ferguson, D.W. Hough, M.J. Danson, and G.L. Taylor. 1997. The crystal structure of citrate synthase from the hyperthermophilic archaeon *Pyrococcus furiosus* at 1.9 Å resolution. Biochemistry 36:9983–94.

Saloheimo, M., J. Kuja-Panula, E. Ylösmaki, M. Ward, and M. Penttilä. 2002a. Enzymatic properties and intracellular localization of the novel *Trichoderma reesei* beta-glucosidase BGLII (cel1A). Appl Environ Microbiol 68:4546–53.

Saloheimo, M., M. Paloheimo, S. Hakola, J. Pere, B. Swanson, E. Nyyssönen, A. Bhatia, M. Ward, and M. Penttilä. 2002b. Swollenin, a *Trichoderma reesei* protein with

sequence similarity to the plant expansins, exhibits disruption activity on cellulosic materials. Eur J Biochem 269:4202–4211.

Saloheimo, M., T. Nakari-Setalä, M. Tenkanen, and M. Penttilä. 1997. cDNA cloning of a *Trichoderma reesei* cellulase and demonstration of endoglucanase activity by expression in yeast. Eur J Biochem 249:584–91.

Saloheimo, A., B. Henrissat, A.M. Hoffren, O. Teleman, and M. Penttilä. 1994. A novel, small endoglucanase gene, egl5, from *Trichoderma reesei* isolated by expression in yeast. Mol Microbiol 13:219–28.

Saloheimo, M., P. Lehtovaara, M. Penttila, T.T. Teeri, J. Stahlberg, G. Johansson, G. Pettersson, M. Claeyssens, P. Tomme, and J.K. Knowles. 1988. EGIII, a new endoglucanase from *Trichoderma reesei*: the characterization of both gene and enzyme. Gene 63:11–22.

Sambrook, J., and D.W. Russel. 2001. Molecular cloning. 3rd edition ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Schimmel, P.R., and P.J. Flory. 1968. Conformational energies and configurational statistics of copolypeptides containing L-proline. J Mol Biol 34:105–20.

Schwarz, W.H. 2001. The cellulosome and cellulose degradation by anaerobic bacteria. Appl Microbiol Biot 56:634–649.

Schwede, T., J. Kopp, N. Guex, and M.C. Peitsch. 2003. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res 31:3381–3385.

Schülein, M. 1997. Enzymatic properties of cellulases from *Humicola insolens*. J Biotechnol 57:71–81.

Scigelova, M., P.S. Green, A.E. Giannakopulos, A. Rodger, D.H.G. Crout, and P.J. Derrick. 2001. A practical protocol for the reduction of disulfide bonds in proteins prior to analysis by mass spectrometry. Eur J Mass Spect 7:29–34.

Shen, H., N.R. Gilkes, D.G. Kilburn, R.C. Miller, Jr., and R.A. Warren. 1995. Cellobiohydrolase B, a second exo-cellobiohydrolase from the cellulolytic bacterium Cellulomonas fimi. Biochem J 311 (Pt 1):67–74.

Shih, P., and J.F. Kirsch. 1995. Design and structural analysis of an engineered thermostable chicken lysozyme. Protein Sci 4:2063–72.

Shoseyov, O., Z. Shani, and I. Levy. 2006. Carbohydrate binding modules: biochemical properties and novel applications. Microbiol Mol Biol Rev 70:283–295.

Sinnott, M.L. 1990. Catalytic mechanisms of enzymic glycosyl transfer. Chem Rev 90:1171–1202.

Song, J.K., and J.S. Rhee. 2000. Simultaneous Enhancement of Thermostability and Catalytic Activity of Phospholipase A1 by Evolutionary Molecular Engineering. Appl Environ Microb 66:890–894.

Srisodsuk, M., K. Kleman-Leyer, S. Keränen, T.K. Kirk, and T.T. Teeri. 1998. Modes of action on cotton and bacterial cellulose of a homologous endoglucanase-exoglucanase pair from *Trichoderma reesei*. Eur J Biochem 251:885–92.

Srisodsuk, M., T. Reinikainen, M. Penttilä, and T.T. Teeri. 1993. Role of the interdomain linker peptide of *Trichoderma reesei* cellobiohydrolase I in its interaction with crystalline cellulose. J Biol Chem 268:20756–20761.

Stals, I., K. Sandra, S. Geysens, R. Contreras, J. Van Beeumen, and M. Claeyssens. 2004. Factors influencing glycosylation of *Trichoderma reesei* cellulases. I: Postsecretorial changes of the O- and N-glycosylation pattern of Cel7A. Glycobiology 14:713–24.

Steipe, B., B. Schiller, A. Pluckthun, and S. Steinbacher. 1994. Sequence statistics reliably predict stabilizing mutations in a protein domain. J Mol Biol 240:188–92.

Stemmer, W.P.C. 1994. DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. Proc Natl Acad Sci 91:10747–10751.

Sterner, R., and W. Liebl. 2001. Thermophilic adaptation of proteins. Crit Rev Biochem Mol Biol 36:39–106.

Ståhlberg, J., C. Divne, A. Koivula, K. Piens, M. Claeyssens, T.T. Teeri, and T.A. Jones. 1996. Activity studies and crystal structures of catalytically deficient mutants of cellobiohydrolase I from *Trichoderma reesei.* J Mol Biol 264:337–349.

Ståhlberg, J., G. Johansson, and G. Pettersson. 1991. A new model for enzymatic hydrolysis of cellulose based on the two-domain structure of cellobiohydrolase I. Bio/Technology 9:286–290.

Sugiyama, J., T. Okano, H. Yamamoto, and F. Horii. 1990. Transformation of Valonia cellulose crystals by an alkaline hydrothermal treatment. Macromolecules 23:3196–3198.

Sulzenbacher, G., H. Driguez, B. Henrissat, M. Schulein, and G.J. Davies. 1996. Structure of the *Fusarium oxysporum* endoglucanase I with a nonhydrolyzable substrate analogue: substrate distortion gives rise to the preferred axial orientation for the leaving group. Biochemistry 35:15280–7.

Suurnäkki, A., M. Tenkanen, M. Siika-Aho, M.-L. Niku-Paavola, L. Viikari, and J. Buchert. 2000. *Trichoderma reesei* cellulases and their core domains in the hydrolysis and modification of chemical pulp. Cellulose 7:189–209.

Swers, J.S., B.A. Kellogg, and K.D. Wittrup. 2004. Shuffled antibody libraries created by *in vivo* homologous recombination and yeast surface display. Nucleic Acids Reseach 32:DOI:10.1093/nar/gnh030.

Szilagyi, A., and P. Zavodszky. 2000. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. Structure 8:493–504.

Takashima, S., M. Ohno, M. Hidaka, A. Nakamura, H. Masaki, and T. Uozumi. 2007. Correlation between cellulose binding and activity of cellulose-binding domain mutants of *Humicola grisea* cellobiohydrolase 1. FEBS Lett 581:5891–6.

Takashima, S., A. Nakamura, M. Hidaka, H. Masaki, and T. Uozumi. 1999. Molecular cloning and expression of the novel fungal beta-glucosidase genes from *Humicola grisea* and *Trichoderma reesei*. J Biochem 125:728–736.

Takashima, S., H. Iikura, A. Nakamura, M. Hidaka, H. Masaki, and T. Uozumi. 1998. Overproduction of recombinant *Trichoderma reesei* cellulases by *Aspergillus oryzae* and their enzymatic properties. J Biotechnol 65:163–71.

Tanner, W., and L. Lehle. 1987. Protein glycosylation in yeast. Biochim Biophys Acta 906:81–99.

Teeri, T.T. 1997. Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. Tibtech 15:160–167.

Teeri, T.T., A. Koivula, M. Linder, G. Wohlfahrt, C. Divne, and T.A. Jones. 1998. *Trichoderma reesei* cellobiohydrolases: why so efficient on crystalline cellulose? Biochem Soc Trans 26:173–178.

Tenkanen, M., M. Makkonen, M. Perttula, L. Viikari, and A. Teleman. 1997. Action of *Trichoderma reesei* mannanase on galactoglucomannan in pine kraft pulp. J Biotechnol 57:191–204.

Tomme, P., S.I. McRae, T.M. Wood, and M. Claeyssens. 1988a. Chromatographic separation of cellulolytic enzymes. Methods Enzymol 160:187–192.

Tomme, P., H. Van Tilbeurgh, G. Pettersson, J. Van Damme, J. Vandekerckhove, J. Knowles, T. Teeri, and M. Claeyssens. 1988b. Studies of the cellulolytic system of *Trichoderma reesei* QM 9414. Analysis of domain function in two cellobiohydrolases by limited proteolysis. Eur J Biochem 170:575–81.

Tracewell, C.A., and F.H. Arnold. 2009. Directed enzyme evolution: climbing fitness peaks one amino acid at a time. Curr Opin Chem Biol doi:10.1016/j.cbpa.2009.01.017.

Tsukada, T., K. Igarashi, M. Yoshida, and M. Samejima. 2006. Molecular cloning and characterization of two intracellular beta-glucosidases belonging to glycoside hydrolase family 1 from the basidiomycete *Phanerochaete chrysosporium*. Appl Microbiol Biotechnol 73:807–14.

Tuohy, M.G., D.J. Walsh, P.G. Murray, M. Claeyssens, M.M. Cuffe, A.V. Savage, and M.P. Coughlan. 2002. Kinetic parameters and mode of action of the cellobiohydrolases produced by *Talaromyces emersonii*. Biochim Biophys Acta 1596:366–380.

Walseth, C.S. 1952. Occurrence of cellulases in enzyme preparations from micro-organisms. Tappi (American Technical Association of the Pulp and Paper Industry) 35:228–233.

van den Burg, B., and V.G. Eijsink. 2002. Selection of mutations for increased protein stability. Curr Opin Biotechnol. 13:333–337.

van den Burg, B., G. Vriend, O.R. Veltman, G. Venema, and V.G.H. Eijsink. 1998. Engineering an enzyme to resist boiling. Proc Natl Acad Sci USA 95:2056–2060.

van Tilbeurgh, H., and M. Claeyssens. 1985. Detection and differentiation of cellulase components using low molecular mass fluorogenic substrates. FEBS Lett. 187:283–288.

van Tilbeurgh, H., F.G. Loonties, C.K. de Bruyne, and M. Claeyssens. 1988. Fluorogenic and chromogenic glycosides as substrates and ligands of carbohydrases. Methods Enzymol 160:45–59.

van Tilbeurgh, H., P. Tomme, M. Claeyssens, R. Bhikhabbai, and G. Pettersson. 1986. Limited proteolysis of the cellobiohydrolase I from *Trichoderma reesei*. FEBS Lett 204:223–227.

Wang, C., M. Eufemi, C. Turano, and A. Giartosio. 1996. Influence of the carbohydrate moiety on the stability of glycoproteins. Biochemistry 35:7299–307.

Wang, T., X. Liu, Q. Yu, X. Zhang, Y. Qu, and P. Gao. 2005. Directed evolution for engineering pH profile of endoglucanase III from *Trichoderma reesei*. Biomol Eng 22:89–94.

Watanabe, H., and G. Tokuda. 2001. Animal cellulases. Cell Mol Life Sci 58:1167–1178.

Watanabe, K., T. Masuda, H. Ohashi, H. Mihara, and Y. Suzuki. 1994. Multiple proline substitutions cumulately thermostabilize *Bacillus cereus* ATCC7064 oligo-1,6-glucosidase. Irerefragable proof supporting proline rule. Eur. J. Biochem. 226:277–283.

Vehmaanperä, J., M. Alapuranen, T. Puranen, M. Siika-aho, J. Kallio, S. Hooman, S. Voutilainen, T. Halonen, and L. Viikari. 2007. Treatment of cellulosic material and enzymes useful therein Patent WO2007071818 A1 2007.

Vehmaanperä, J., M. Elovainio, H. Haakana, V. Joutsjoki, R. Lantto, A. Mäntylä, M. Paloheimo, P. Suominen, J. Londesborough, and A. Miettinen-Oinonen. 1997. Cellulases, the genes encoding them and uses thereof. International patent application Patent WO97/14804 1997.

Veronese, F.M., E. Boccu, O. Schiavon, C. Grandi, and A. Fontana. 1984. General stability of thermophilic enzymes: studies on 6-phosphogluconate dehydrogenase from *Bacillus stearothermophilus* and yeast. J Appl Biochem 6:39–47.

Vieille, C., and G.J. Zeikus. 2001. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. Microbiol Mol Biol Rev 65:1–43.

Vieille, C., K.L. Epting, R.M. Kelly, and J.G. Zeikus. 2001. Bivalent cations and amino-acid composition contribute to the thermostability of *Bacillus licheniformis* xylose isomerase. Eur J Biochem 268:6291–301.

Vieille, C., J.M. Hess, R.M. Kelly, and J.G. Zeikus. 1995. xylA cloning and sequencing and biochemical characterization of xylose isomerase from Thermotoga neapolitana. Appl Environ Microbiol 61:1867–75.

Viikari, L., M. Alapuranen, T. Puranen, J. Vehmaanpera, and M. Siika-Aho. 2007. Thermostable enzymes in lignocellulose hydrolysis. Adv Biochem Eng Biotechnol 108:121–45.

Vuong, T.V., and D.B. Wilson. 2010. Glycoside hydrolases: Catalytic Base/Nuclephile Diversity. Biotech Bioeng 107:195–205.

Williams, J.C., J.P. Zeelen, G. Neubauer, G. Vriend, J. Backmann, P.A. Michels, A.M. Lambeir, and R.K. Wierenga. 1999. Structural and mutagenesis studies of leishmania triosephosphate isomerase: a point mutation can convert a mesophilic enzyme into a superstable enzyme without losing catalytic power. Protein Eng 12:243–50.

Wilson, D.B. 2008. Three microbial strategies for plant cell wall degradation. Ann N Y Acad Sci 1125:289–97.

Wilson, D.B. 2004. Studies of Thermobifida fusca plant cell wall degrading enzymes. Chem Rec 4:72–82.

Withers, S.G. 2001. Mechanisms of glycosyl transferases and hydrolases. Carbohyd Polym 44:325–337.

Vogt, G., S. Woell, and P. Argos. 1997. Protein thermal stability, hydrogen bonds, and ion pairs. J Mol Biol 269:631–43.

Wohlfahrt, G., T. Pellikka, H. Boer, T.T. Teeri, and A. Koivula. 2003. Probing pH-dependent functional elements in proteins: modification of carboxylic acid pairs in *Trichoderma reesei* cellobiohydrolase Cel6A. Biochemistry 42:10095–103.

Voigt, C.A., C. Martinez, Z.G. Wang, S.L. Mayo, and F.H. Arnold. 2002. Protein building blocks preserved by recombination. Nat Struct Biol 9:553–8.

von Ossowski, I., T. Teeri, N. Kalkkinen, and C. Oker-Blom. 1997. Expression of a fungal cellobiohydrolase in insect cells. Biochem Biophys Res Commun 233:25–9.

Wood, T.M., and K.M. Bhat. 1988. Methods for measuring cellulase activities. Methods in enzymology 160:87–112.

Wrba, A., A. Schweiger, V. Schultes, R. Jaenicke, and P. Zavodszky. 1990. Extremely thermostable D-glyceraldehyde-3-phosphate dehydrogenase from the eubacterium *Thermotoga maritima*. Biochemistry 29:7584–92.

Wymelenberg, A.V., S. Denman, D. Dietrich, J. Bassett, X. Yu, R. Atalla, P. Predki, U. Rudsander, T.T. Teeri, and D. Cullen. 2002. Transcript analysis of genes encoding a family 61 endoglucanase and a putative membrane-anchored family 9 glycosyl hydrolase from Phanerochaete chrysosporium. Appl Environ Microbiol 68:5765–8.

Xiong, H., F. Fenel, M. Leisola, and O. Turunen. 2004. Engineering the thermostability of *Trichoderma reesei* endo-1,4-beta-xylanase II by combination of disulphide bridges. Extremophiles 8:393–400.

Yamamoto, H., F. Horii, and H. Odani. 1989. Structural changes of native cellulose crystals induced by annealing in aqueous alkaline and acidic solutions at high temperatures. Macromolecules 22:4130–4132.

Yan, A., and W.J. Lennarz. 2005. Unraveling the mechanism of protein N-glycosylation. J Biol Chem 280:3121–4.

Yuan, J.S., K.H. Tiller, H. Al-Ahmad, N.R. Stewart, and C.N. Stewart, Jr. 2008. Plants to power: bioenergy to fuel the future. Trends Plant Sci 13:421–9.

Zaccolo, M., D.M. Williams, D.M. Brown, and E. Gherardi. 1996. An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. J Mol Biol. 225:589–603.

Zhang, J.-H., G. Dawes, and W.P.C. Stemmer. 1997. Directed evolution of a fucosidase from a galactosidase by DNA shuffling and screening. Proc Natl Acad Sci 94:4504–4509.

Zhang, S., D.C. Irwin, and D.B. Wilson. 2000. Site-directed mutation of noncatalytic residues of Thermobifida fusca exocellulase Cel6B. Eur J Biochem 267:3101–15.

Zhang, X.J., W.A. Baase, B.K. Shoichet, K.P. Wilson, and B.W. Matthews. 1995. Enhancement of protein stability by the combination of point mutations in T4 lysozyme is additive. Protein Engin. 8:1017–1022.

Zhang, Y.H., and L.R. Lynd. 2006. Biosynthesis of radiolabeled cellodextrins by the *Clostridium thermocellum* cellobiose and cellodextrin phosphorylases for measurement of intracellular sugars. Appl Microbiol Biotechnol 70:123–9.

Zhang, Y.H., and L.R. Lynd. 2004. Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems. Biotechnol Bioeng 88:797–824.

Zhang, Y.-H.P., M.E. Himmel, and J.R. Mielenz. 2006. Outlook for cellulase improvement: Screening and selection strategies. Biotechnol Adv 24:452–481.

Zhao, H., L. Giver, Z. Shao, J.A. Affholter, and F.H. Arnold. 1998. Molecular evolution by staggered extension process (StEP) in vitro recombination. Nat Biotechnol 16:258–261.

Zou, J., G.J. Kleywegt, J. Stahlberg, H. Driguez, W. Nerinckx, M. Claeyssens, A. Koivula, T.T. Teeri, and T.A. Jones. 1999. Crystallographic evidence for substrate ring distortion and protein conformational changes during catalysis in cellobiohydrolase Cel6A from *Trichoderma reesei*. Structure 7:1035–45.

Publications I–IV are not included in the PDF version.
Please order the printed version to get the complete publication
(http://www.vtt.fi/publications/index.jsp).

Series title, number and report code of publication

VTT Publications 754
VTT-PUBS-754

Author(s)
Sanni Voutilainen

Title

Fungal thermostable cellobiohydrolases
Characterization and protein engineering studies

Abstract

Cellulases are important industrial enzymes, and they are already today utilized for example in the pulp and paper and textile industries. Due to their application potential in so-called second generation bioethanol production, starting from lignocellulosic raw materials, cellulases are currently in the centre of attention. Especially cellobiohydrolases are the key enzymes in total hydrolysis of biomass based processes. The currently known cellobiohydrolases from GH family 7 are, however, not highly active under industrial conditions.

In this work two different approaches were taken to improve the hydrolysis of crystalline cellulose. Firstly, thermotolerance and activity of two known fungal cellobiohydrolases from GH7 family was improved by protein engineering and secondly novel cellobiohydrolases were studied.

Engineering of GH7 cellobiohydrolases has been hindered because of difficulties in heterologous expression in a bacterial or yeast host. In this study a functional yeast expression system was developed for two single-module cellobiohydrolases. This heterologous expression system enabled protein engineering by random mutagenesis and site-directed mutagenesis. Random mutagenesis, carried out with error-prone PCR and followed by functional screening with an automated, thermostability screening method, was used for improving the thermostability of *Melanocarpus albomyces* Cel7B (*Ma* Cel7B). The stability and activity of *Ma* Cel7B was further improved through structure-guided protein engineering by introducing an additional disulphide bridge and a carbohydrate binding module. Rational mutagenesis was also used for engineering *Talaromyces emersonii* Cel7A (*Te* Cel7A). Altogether five individual S-S bridges were introduced to improve the stability of *Te* Cel7A. Three out of these five single S-S mutants had a clearly improved thermostability. These mutations were combined in a triple mutant, which had significantly improved unfolding temperature (by 9°C) and ability to hydrolyse microcrystalline cellulose at 80°C.

In addition to the mutagenesis studies of known cellobiohydrolases, we found novel GH-7 family cellobiohydrolases, which have high activity on insoluble polymeric substrates. Three family 7 cellobiohydrolases originating from thermophilic fungi *Acremonium thermophilum*, *Thermoascus aurantiacus* and *Chaetomium thermophilum* were characterized and compared to the widely studied *Trichoderma reesei* Cel7A enzyme. The comparison revealed that all these novel cellobiohydrolases are promising for application purposes, because they were more thermostable than *T. reesei* Cel7A and more active in the hydrolysis of microcrystalline cellulose (Avicel) at 45°C.

ISBN
978-951-38-7425-4 (soft back ed.)
978-951-38-7426-1 (URL: http://www.vtt.fi/publications/index.jsp)

Series title and ISSN
VTT Publications
1235-0621 (soft back ed.)
1455-0849 (URL: http://www.vtt.fi/publications/index.jsp)

Project number

Date
December 2010

Language
English

Pages
98 p. + app. 55 p.

Name of project

Commissioned by

Keywords
Cellulase, protein engineering/, thermostability, heterologous expression, *Saccharomyces cerevisiae*, high-throughput screening, random mutagenesis, site-directed mutagenesis, disulphide bridge

Publisher
VTT Technical Research Centre of Finland
P.O. Box 1000, FI-02044 VTT, Finland
Phone internat. +358 20 722 4520
Fax +358 20 722 4374

VTT PUBLICATIONS

ISBN 978-951-38-7425-4 (soft back ed.)
ISSN 1235-0621 (soft back ed.)

ISBN 978-951-38-7426-1 (URL: http://www.vtt.fi/publications/index.jsp)
ISSN 1455-0849 (URL: http://www.vtt.fi/publications/index.jsp)